DOCUMENT RESUME

ED 468 800 SE 064 476

AUTHOR Atkins, John F.; Ellington, Andrew; Friedman, B. Ellen; Gesteland, Raymond F.; Noller, Harry F.; Pasquale, Stephen M.; Storey, Richard D.; Uhlenbeck, Olke C.; Weiner, Alan M.

TITLE Bringing RNA into View: RNA and Its Roles in Biology.

INSTITUTION Biological Sciences Curriculum Study, Colorado Springs.

SPONS AGENCY National Science Foundation, Arlington, VA.

PUB DATE 2000-00-00

NOTE 194p.

CONTRACT NSF-9652921

AVAILABLE FROM BSCS, Pikes Peak Research Park, 5415 Mark Dabling Blvd., Colorado Springs, CO 80918-3842. Tel: 719-531-5550; Web site: http://www.bscs.org.

PUB TYPE Guides - Classroom - Learner (051) -- Guides - Classroom - Teacher (052)

EDRS PRICE EDRS Price MF01/PC08 Plus Postage.

DESCRIPTORS *Science Activities; Biology; *Genetics; Higher Education; *Instructional Materials; *RNA; Science Instruction

ABSTRACT

This guide presents a module for college students on ribonucleic acid (RNA) and its role in biology. The module aims to integrate the latest research and its findings into college-level biology and provide an opportunity for students to understand biological processes. Four activities are presented: (1) "RNA Structure: Tapes to Shapes"; (2) "RNA Catalysis"; (3) "RNA and Evolution"; and (4) "RNA Evolution in Health and Disease." (Contains 28 references.) (YDS)

Reproductions supplied by EDRS are the best that can be made from the original document.

Bringing RNA into View

RNA and Its Roles in Biology

ED 468 800

BEST COPY AVAILABLE

SE064476

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

b. Perrin

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

2

BSCS

Bringing RNA into View: RNA and Its Roles in Biology

Biological Sciences Curriculum Study (BSCS)
Pikes Peak Research Park
5415 Mark Dabling Blvd.
Colorado Springs, Colorado 80918-3842

Authors

John E. Atkins, *Eccles Institute of Human Genetics, University of Utah*
Andrew Ellington, *Indiana University*
B. Ellen Friedman, *Consultant*
Raymond F. Gesteland, *Eccles Institute of Human Genetics, University of Utah*
Harry F. Noller, *University of California—Santa Cruz*
Stephen M. Pasquale, *BSCS*
Richard D. Storey, *The Colorado College*
Olke C. Uhlenbeck, *University of Colorado, Boulder*
Alan M. Weiner, *Yale University*

Copyright © 2000 by the BSCS. All rights reserved. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage or retrieval system, without permission in writing. For permissions and other rights under this copyright, please contact the BSCS, Pikes Peak Research Park, 5415 Mark Dabling Blvd., Colorado Springs, CO 80918-3842.

This material is based on work supported by the National Science Foundation under grant No. 9652921. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the granting agency.

BSCS Development Team
Project Leadership
Joseph D. McInerney, *Principal Investigator*
Stephen M. Pasquale, *Project Director*

Project Coordination
Rose Johnson, *Production Assistant*
Dee Miller, *Production Assistant*
Barbara C. Resch, *Editor*

Project Support
Linda Ward, *Senior Executive Assistant*
Kevin Andrews, *Graphic Artist*
Lisa M. Chilberg, *Graphic Designer*

BSCS Administrative Staff
Timothy H. Goldsmith, *Chair, Board of Directors*
Rodger W. Bybee, *Executive Director*
Susan Loucks-Horsley, *Associate Executive Director*
Marcia Mitchell, *Financial Officer*

Field-Test Teachers
David Allard, *Texarkana College & Texas A&M University-Texarkana, Texas*
Gail Carmack, *The Science Academy of Austin at LBJ High School, Texas*
Alix Darden, *The Citadel, South Carolina*
Denise M. Foley, *Chapman University, California*
David A. McCullough, *Wartburg College, Iowa*
John Ohlsson, *University of Colorado, Boulder*
LeLeng P. To, *Goucher College, Maryland*

Design and layout by Angela Greenwalt

Illustration Credits
Figure 1.4a: © A. Barrington Brown, Science Source/Photo Researchers; Figure 1.4b: Courtesy of Harry F. Noller; Figure 1.4c: *Genes and Genomes* by Maxine Singer and Paul Berg, published by University Science Books, Sausolito, CA; Figure 2.3: John T. Ohlsson, University of Colorado, Boulder; Figure 3.6: stopwatch © 1999–2000 www.barrysclipart.com; Figures 3.9, T3.1, and T3.2: Gerald F. Joyce and Kathleen E. McGinness, The Scripps Research Institute; Figure 4.1b: Noller, H.F. (1993). On the origin of the ribosome. In Gesteland, R.F., & Atkins, J.F. (Eds.), *The RNA world* (p. 141). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Figures 8 and 1.6 are adapted from Gesteland, R.F., & Atkins, J.F. (1993). *The RNA world.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Evaluation Form for
Bringing RNA into View: RNA and Its Roles in Biology

Your feedback is important. After you have used the module, please take a few minutes and return this form to BSCS, Attn: RNA, 5415 Mark Dabling Blvd., Colorado Springs, CO 80918-3842.

1. Please evaluate the *Faculty Background* by marking this form and providing written comments or suggestions on a separate sheet.

Sections Used	not helpful				very helpful
Overview	1	2	3	4	5
Background on RNA	1	2	3	4	5

2. Please evaluate the activities by marking this form and providing written comments or suggestions on a separate sheet. Rate the activities for their effectiveness at teaching concepts of RNA and its role in biology.

Activity 1: RNA Structure: Tapes to Shapes	1	2	3	4	5
Activity 2: RNA Catalysis	1	2	3	4	5
Activity 3: RNA and Evolution	1	2	3	4	5
Activity 4: RNA Evolution in Health and Disease	1	2	3	4	5

3. What are the major strengths of this module?

4. What are the major weaknesses of this module?

5. Please rate the overall effectiveness of this module: not effective very effective

1 2 3 4 5

6. Please provide a description of the classes in which you used this module: (circle response)

College: 2 year 4 year High school: grade 9 10 11 12

freshman sophomore junior senior Level of class: basic honors 2nd year

How many students used the module? _____ How many students per class? _____

Ethnicity (approximate % of minorities): _____

Description of school: _____

College: liberal arts science High school: urban suburban rural

7. Have you used BSCS materials before? ❒ yes ❒ no

8. Please provide your name and contact information below:

Name_____

School_____

Mailing address_____ ❒ home ❒ work

_____

_____

Phone_____ ❒ home ❒ work

FAX_____ ❒ home ❒ work

E-mail address_____

Was the address on your mailing label correct? ❒ yes ❒ no

Contents

Faculty Background

Overview . . . 3
Background on RNA . . . 5
References and Related Literature . . . 19

Annotated Faculty Pages

About the Activities . . . 23
Activity 1 RNA Structure: Tapes to Shapes . . . 25
Activity 2 RNA Catalysis . . . 45
Activity 3 RNA and Evolution . . . 65
Activity 4 RNA Evolution in Health and Disease . . . 95

Copymasters . . . 115

Templates . . . 123

Student Pages

Activity 1 RNA Structure: Tapes to Shapes . . . 133
Activity 2 RNA Catalysis . . . 145
Activity 3 RNA and Evolution . . . 161
Activity 4 RNA Evolution in Health and Disease . . . 181

Faculty Background

BEST COPY AVAILABLE

Overview

Introduction

The impetus for developing *Bringing RNA into View* for college biology classes is the recent and rapid growth in knowledge of the structures and diverse functions of RNA molecules. First described in 1947 as a cellular constituent involved in protein synthesis, RNA has since been shown to play several other essential roles in gene expression, including genome maintenance, processing and editing of primary transcripts, and localization of proteins within the cell. New RNAs continue to be discovered performing unexpected tasks in the cell.

Perhaps more than any other single discovery about nucleic acids since Watson and Crick's elucidation of the DNA double helix, the finding that RNA (and DNA) has catalytic ability has expanded our view of these molecules' potential, as evolutionary progenitors and contemporary biochemical players. The notion that an RNA world phase occurred early in prebiotic evolution (deriving both its information encoding and catalytic functions from RNA molecules) has inspired a burst of research into the molecular origins of life and the biochemical potential of nucleic acids.

Particularly revealing of the impressive biochemical potential of these molecules are recent studies expanding on the 30-year-old observation of Darwinian selection and evolution of nucleic acids in the test tube. Technological advances of the last 15 years have greatly increased the power and versatility of such *in vitro* experiments to create and select talented nucleic acid molecules, some of which have begun to reveal RNA's potential for true self-replication, for synthesizing its own monomer building blocks, and for practical use in combating viral and microbial infections as well as genetic disorders.

BSCS selected the topic of this module, RNA and its role in biology, as a key area for synthesizing these important new research findings into the fundamental concepts of college-level biology. In addition, this topic offers a useful opportunity to shift students from focusing on isolated facts to approaching biology conceptually; in short, the module helps students think about biological processes.

To develop this module, biologists at the Biological Sciences Curriculum Study in Colorado Springs and the Eccles Institute of Human Genetics at the University of Utah in Salt Lake City worked with an external advisory committee of scientists and educators, plus a variety of college faculty who conducted field tests of the module. This process identified the following major concepts for the module:

- Nucleic acids (DNA and particularly RNA) have two major functions: as informational molecules and as biochemical catalysts.
- The sequence of monomers in RNA dictates its three-dimensional structure and, consequently, its function.
- Molecules can be subject to natural selection.
- Evolution requires an iterative process of molecular replication, mutation, and selection.
- Modern roles for RNA suggest its major role in the origin of life.
- Studies of the origin of life and its evolution work with substantial data, such as modern observation of naturally occurring ribozymes (RNA catalysts) and *in vitro* molecular selection experiments.

- Modern roles for RNA include serving as a genome to a variety of viruses and smaller viroids.
- Most RNA-based viruses and viroids are significant pathogens of humans, animals, and agriculturally important crop plants, and as such they have a major social and economic impact.

How to Use the Module

The module provides background materials for faculty and a set of four educational activities for students. *Background on RNA* (in this section) provides you with an update on RNA research in a form that is accessible to busy faculty. This material is for your own use. It may extend your knowledge and thus be helpful for teaching the activities, but it is not essential to teaching the activities. The activities are inquiry-based explorations that offer you an alternative to lecture and stimulate student interest and responsibility for learning.

Figure 1 shows the layout of the materials. The module features four core classroom and laboratory activities that explore RNA's structure and function, RNA catalysis, RNA replication and evolution, and RNA's role in health and disease.

Notice that each activity appears in two sections, *Annotated Faculty Pages* and *Student Pages*. The *Student Pages* consist of introductory text and For Your Information essays that provide context and elaboration for the activity. Detailed protocols are provided for students, and Challenge Questions stimulate student thought and synthesis of ideas. The *Student Pages* and the *Copymasters* and *Templates* may be photocopied for classroom use.

Figure 1 The module at a glance. *Bringing RNA into View* consists of three main components (*Faculty Background*, *Annotated Faculty Pages*, and *Student Pages*) as well as support materials.

Faculty Background
- Overview
- Background on RNA
- References

Annotated Faculty Pages
- Instructions and background for Activities 1–4

Copymasters and Templates
- Handouts and masters for Activities 1–4

Student Pages
- Student materials for Activities 1–4

The *Annotated Faculty Pages* contain the student text (**bold type**) plus annotations for the faculty (regular type). The annotations contain suggestions and hints for teaching each activity, answers to Challenge Questions, optional extension exercises, and reagent preparation instructions.

A summary of the activities is provided in Figure 2. Time and resources may not permit you to teach all four of the activities. However, we recommend that you teach Activity 1 (*RNA Structure: Tapes to Shapes*) before the other activities, to ensure that students understand the structural basis for the RNA functions explored in the later activities.

As your students proceed through the module, we recommend that you encourage them to ask questions, seek outside resources, and be aware of the way in which science attempts to understand natural processes. For example, call students' attention to citations in essays so that they begin to appreciate the significance of discussions that are based on primary scientific data rather than hearsay.

Figure 2 Summary of the student activities.

Activity 1
RNA Structure: Tapes to Shapes
Students apply rules of base pairing and folding to construct physical models of RNA sequences. They use their models to explore structure-function relationships and the effects of mutation.

Activity 2
RNA Catalysis
Students explore catalytic RNA in the laboratory using a self-splicing group I intron. Students apply the techniques of *in vitro* transcription, RNA isolation, and acrylamide electrophoresis to study the kinetics of the splicing reaction.

Activity 3
RNA and Evolution
Students explore the replication of a catalytic RNA in the laboratory using a continuous *in vitro* system.

Activity 4
RNA Evolution in Health and Disease
Students explore the continuing evolution of RNA in the context of the emerging resistance of bacteria and viruses to therapeutic agents.

Background on RNA

The Path to the RNA World Hypothesis

The classic formulation of the flow of genetic information during gene expression holds that DNA is copied both to itself and to RNA, and RNA is then decoded to synthesize protein. This view was formalized by Francis Crick in 1968 as the central dogma of molecular biology (Figure 3).

Several assumptions are inherent in this traditional view: Information flow is unidirectional, as expressed by the one-way arrows; only DNA is a template that can be replicated; nucleic acid coding information must ultimately be translated to protein form if working catalysts are to result; and DNA, as the ultimate molecule that encodes genetic information, likely preceded RNA during the formation and subsequent evolution of biomolecules.

This straightforward view of gene expression has given way to a more detailed understanding of information flow in biology. Molecular genetic research, particularly during the last 15 years, provides exciting new insights that reveal the original formulation of the central dogma to be incomplete. Some of the assumptions listed above have been shown to be limiting. For example, the study of viruses having an RNA genome revealed that RNA, like DNA, can serve as a primary information-encoding molecule. Also, the discovery in 1970 of reverse transcriptase, an RNA virus-encoded protein catalyst that copies RNA-based information into DNA form, revealed that information flow in the biological world is in fact a two-way street between DNA and RNA.

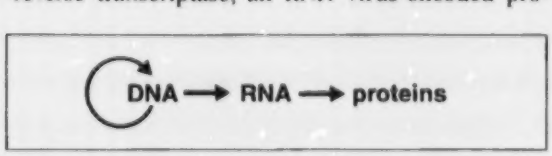


Figure 3 Central dogma of molecular biology.

Watson and Crick appreciated DNA's potential to be a template for its own replication as soon as they solved the double helical structure of the molecule. We now realize that RNA molecules likewise can play template roles for their own replication. In the process known as *template-directed nucleic acid replication*, an RNA sequence can serve as an informative scaffold onto which complementary bases align to produce a complementary strand (Figure 4a). Many viruses infecting plants and animals, such as plant viroids and polio virus, employ a two-step replication strategy in which their single-stranded RNA genome, acting as a template, is initially transcribed by a replicase enzyme to yield a complementary RNA strand. This molecule in turn serves as a template whose transcription regenerates copies of the original genome strand, thereby effecting replication (Figure 4b).

The discovery of RNA's informational and template functions expanded our ideas of the early evolution of information storage and expression mechanisms. However, the most dramatic and unexpected discovery to influence these ideas came early in the 1980s, when scientists in two different laboratories independently discovered that RNA has catalytic ability. Thomas Cech and his group at the University of Colorado were studying the splicing of large ribosomal RNA (rRNA) precursors in the protozoon *Tetrahymena*. These scientists serendipitously observed that the RNA precursor spontaneously

BEST COPY AVAILABLE

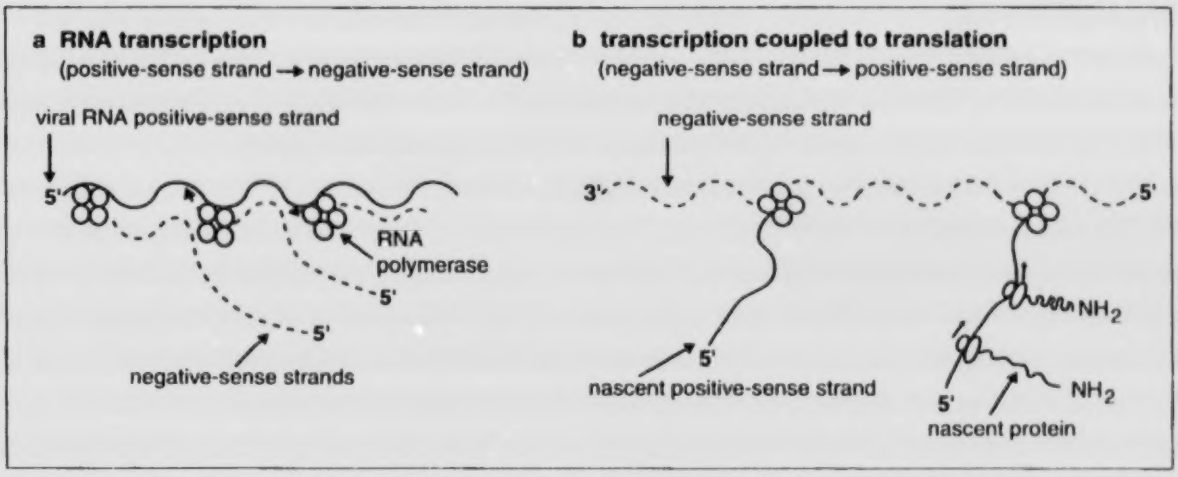


Figure 4 Viral RNA replication. The replication cycle of viruses typically consists of two steps. In (a), the positive-sense RNA genome is transcribed into many copies of a complementary negative-sense strand by RNA polymerase. In (b), the newly formed complementary strands become templates for synthesizing many copies of the original positive-sense genome. The new genomic strands may be translated as they are being synthesized.

changed size, becoming smaller after incubating in a protein-free buffer solution containing only Mg^{2+} ions. Realizing how unusual this result was, they pursued it in earnest. In a series of brilliant experiments, Cech and his coworkers proved that this RNA, the "group I intron," has the inherent ability to catalyze its own excision from the RNA precursor. This RNA intron catalyzes its own self-splicing without the aid of any protein.

Meanwhile, in the lab of Sidney Altman at Yale University, researchers were continuing a long series of biochemical experiments to characterize an enzyme activity in *Escherichia coli*, called *RNase P*, that trimmed the 5'-end of a transfer RNA (tRNA) precursor. After exhaustive purification, the active enzyme was found to contain both a protein component *and* an RNA molecule. The conventional prejudice was that the protein must act as the "enzyme." But any effort to remove the RNA component eliminated the catalytic ability. Again, pursuit of an unexpected finding and diligent experimentation showed that the RNA molecule itself was sufficient to catalyze the trimming reaction. The RNA, in this case, was a catalyst acting not on itself but on another RNA.

Here were two different RNA molecules that catalyzed biochemical reactions, just like proteins. Nobel Prizes followed for both Cech and Altman—a new era had begun. Since this discovery, seven naturally occurring classes of ribozyme have been recognized, and hundreds of specific examples have been identified in a wide variety of organisms (Figure 5). The true prevalence of ribozymes in contemporary cells is unclear, but it is quite likely that new examples will be discovered.

Interestingly, the independent discovery of RNA catalysis by Cech and Altman occurred in research fields that were relative backwaters at the time. In the early 1980s, the hot research fever was in the first thrust of exploiting the new molecular biology technologies (for example, restriction enzymes, cloning, and sequencing) to explore emerging topics such as split genes, tumor viruses, and oncogenes. The unlikely discovery of catalytic RNAs in this setting is testimony to the importance in science of observing and attending to unexpected and unusual results—and being willing to pursue them despite prevailing fashions.

The RNA World Hypothesis

In addition to destroying the orthodox assumption that only proteins function as biological catalysts, the discovery of RNA's chemical versatility led to a dramatic change in how scientists view the likely sequence of molecular events during early evolution. Taken together, RNA's informational, template, and catalytic abilities led to the hypothesis that

Figure 5 Naturally occurring ribozymes.

Class	Size	Reaction	Source
group I intron	large: 413 NT in *Tetrahymena thermophilia*	intron excision	eukaryotes, eubacteria, and viruses
group II intron	large: 887 NT in yeast mitochondria	intron excision	eukaryotic organelles and eubacteria
RNase P	large: 350–410 NT	hydrolytic endoribonuclease	RNA subunit of eubacterial RNase P
hammerhead	small: 31–42 NT (enzyme strand can be 16 NT)	RNA cleavage	viral satellite RNA in plants, viroids, and newt satellite DNA
hairpin	small: 50 NT (minimum sequence)	RNA cleavage	(-) strand satellite RNA of tobacco ringspot virus
hepatitis Delta virus (HDV)	84 NT (required)	RNA cleavage	HDV
Neurospora VS RNA	881 NT (164 sufficient)	RNA cleavage	*Neurospora* mitochondria

RNA evolved, before the appearance of DNA or protein, in an *RNA world*. During this proposed phase of evolution, RNA is assumed to have provided both the coding and the catalytic abilities necessary and sufficient to initiate biological evolution. Specifically, if RNA's catalytic abilities during this time extended to its own *self-replication*, then molecular evolution automatically would have started as randomly variant RNAs were naturally selected on the basis of evermore efficient replication ability. This property of unaided self-replication, although an essential assumption of the RNA world hypothesis, remains to be demonstrated. No nucleic acid sequence possessing RNA replicase catalytic activity has yet been found in nature, but researchers have taken the first steps toward creating such a self-replicator RNA molecule *in vitro*. In a later section, we examine powerful new *in vitro* technologies for directing the evolution of nucleic acid molecules toward this and many other functions, and we discuss their implications for research in both evolutionary biology and biomedical science.

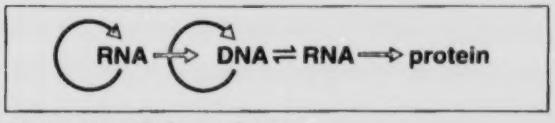


Figure 6 Modified central dogma.

The coding, template, and catalytic abilities of RNA also led to the current, expanded formulation of the central dogma, one that incorporates the RNA world hypothesis and places information flow in an evolutionary, historical context (Figure 6).

The discovery of ribozymes clearly was the catalyst for the current intense interest in the RNA world hypothesis. But the idea that RNA perhaps was the first genetic molecule is not new: Francis Crick and Leslie Orgel first proposed the possibility in 1968. Indeed, today's RNA world hypothesis builds upon a long history of research findings related to the origin of life and molecular evolution (Figure 7).

Next, we discuss RNA shapes and the growing appreciation of their complexity and contribution to the diverse functions of RNA.

RNA Structure: Tapes to Shapes

A key principle of molecular structure is that shape determines function. For biological polymers composed of multiple subunits, the fundamental determinant of shape is the linear information represented in the sequence of individual subunits; the molecule is in effect an information tape. A specific three-dimensional shape emerges in the molecule as thermodynamically favored physical

BEST COPY AVAILABLE

Figure 7 Milestones in the evolution of the RNA world hypothesis.

1800s	The idea of spontaneous generation, belief in the ongoing creation of living organisms from nonliving materials, persists.
1828	Friedrich Wöhler synthesizes urea in the laboratory, eliminating "vital-force" as an agent in the synthesis of organic chemicals.
1859	Charles Darwin publishes *On the Origin of Species*, in which he proposes a theory of biological evolution based on the mechanism of natural selection.
1864	Louis Pasteur experimentally disproves ongoing spontaneous generation by showing that when liquids are boiled in order to kill any microorganisms present, and are subsequently kept sterile, no organisms appear in them.
1924	Alexander Oparin employs geological evidence in proposing that the early earth had a reducing atmosphere lacking oxygen and that the first single-celled organisms might have arisen from simple organic molecules present in this early atmosphere: in effect a restricted version of spontaneous generation.
1929	Biochemist J.B.S. Haldane proposes that life might have arisen on earth when Oparin's early atmosphere was subjected to energy in the form ultraviolet radiation and heat from the cooling earth.
1953	Graduate student Stanley Miller provides experimental support for the Oparin-Haldane hypothesis by mixing gases of the "primitive atmosphere" in a glass reaction vessel and subjecting them to electric current for one week; amino acids are formed *de novo*.
	Biologist James Watson and physicist Francis Crick publish their findings on the structure of DNA.
1961	Marshall Nirenberg and his colleagues begin their five-year project of cracking the genetic code by discovering that a messenger RNA made up entirely of the base uracil can be translated into a peptide made up entirely of the amino acid phenylalanine.
1962	Watson and Crick share a Nobel Prize for their work on DNA structure.
1967	Sol Spiegelman demonstrates the replication and evolution of RNA molecules in the test tube.
1968	Francis Crick and Leslie Orgel propose that the first information molecule was RNA; Crick advances the central dogma of molecular biology.
1970	David Baltimore and Howard Temin independently discover reverse transcription of viral RNA genomes into DNA.
1972	Harry Noller proposes a role for ribosomal RNA in the translation of messenger RNA into protein.
1982–83	Thomas Cech and Sidney Altman independently discover the first examples of catalytic RNA molecules: ribozymes.
1986	Walter Gilbert coins the term "RNA world" to describe the hypothesized time during which RNA was the primary informational and catalytic molecule.
	Kary Mullis develops the polymerase chain reaction (PCR) technology that allows rapid copying of DNA and RNA sequences *in vitro* and enables large-scale laboratory studies of molecular evolution.
1989	Cech and Altman share a Nobel Prize for their discovery of catalytic RNA.
	Gerald Joyce develops the technique of *in vitro* amplification and selection of RNA (that is, directed evolution) using the PCR technique.
1992	Noller presents evidence for the catalytic involvement of the 23S rRNA in peptide bond formation.
1993	Mullis receives a Nobel Prize for his development of the polymerase chain reaction.
	Joyce further develops *in vitro* RNA amplification and evolution experimental procedures.
1995	Jack Szostak's laboratory takes the first steps toward the *in vitro* selection of a self-replicating RNA molecule.
1998	David Bartel and Peter Unrau use *in vitro* selection to demonstrate that RNA can catalyze the formation of individual nucleotides.

interactions, typically noncovalent, occur between compatible subunits located at a distance from one another. Indeed, this important structure-function principle provides the entire rationale for evolution's invention of a genetic information-coding strategy almost 4 billion years ago.

The importance of linear information for molecular shape and function was first appreciated for proteins, the first biopolymers to be sequenced. Their covalently linked amino acid subunits are now known to interact further through a variety of weaker noncovalent associations; these include van der Waals forces, hydrogen bonds, ionic bonds, and hydrophilic interactions. These noncovalent interactions of amino acids, both locally and with more remote neighbors through folding of the molecule, give rise to higher-order protein shapes. More basic, local shape elements, such as hydrogen bond-stabilized alpha-helices and beta-sheets, can interact in a variety of ways within the folded protein. These higher-order, three-dimensional interactions are typically stabilized by hydrophobic and ionic forces to create a vast array of specific protein shapes. We recognize a variety of functional sites in proteins that result from their shapes: enzymatic active sites, binding pockets, regulatory sites, and domains for protein-protein interaction.

The science community's recognition of diversity of shape among nucleic acids developed more slowly, however. DNA's extended double helix, the first nucleic acid structure to be revealed, gave no hint of more complicated shapes. Only later, when the base sequence and three-dimensional structure of transfer RNA (tRNA) was worked out, was the ability of nucleic acids to adopt complex shapes confirmed.

The most important determinant of folding and shape in single-stranded nucleic acids, both RNA and DNA, is complementary base pairing via hydrogen bonding, according to the base-pair rules first established by Watson and Crick. For RNA, pairing of A with U and of G with C is the primary basis for folding. Even though each RNA molecule in a cell normally consists of a single continuous strand, RNA molecules frequently contain linear runs of bases that are complementary to other runs located elsewhere in the molecule. This allows the molecule to fold back and form double-stranded regions within itself.

These double-stranded regions of the molecule adopt a helical configuration similar to that found in double-stranded DNA. (A subtle but characteristic difference, however, is that helical regions in RNA adopt the A-form geometry, whereas those in DNA are most often B-form.) The helical regions alternate with more flexible single-stranded regions. Activity 1 demonstrates that even an RNA molecule of modest length can fold in several possible ways by bringing together different, more or less complementary regions. The degree of match and resulting thermodynamic stability of one structure over an alternative determines which form predominates in the cell.

Until recently, it was difficult to determine the folded structures of RNAs, and only a few were known. Textbooks typically show tRNA as a folded structure,

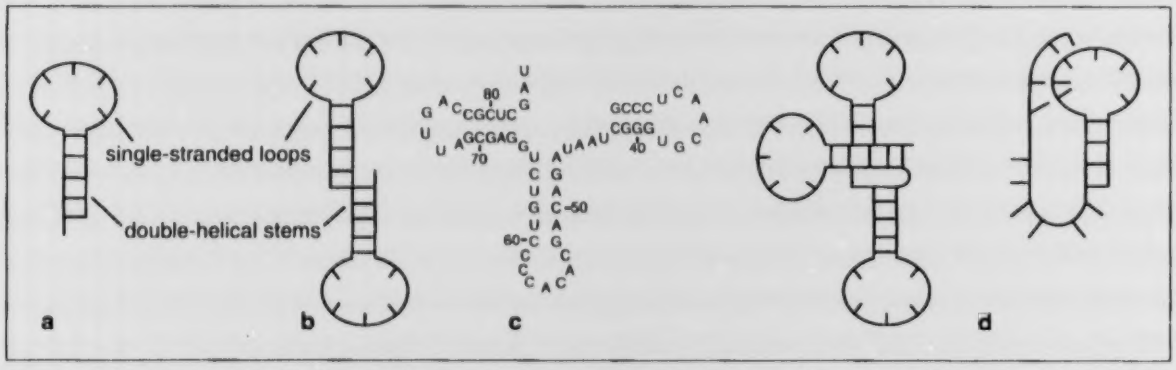


Figure 8 Some examples of RNA shapes and structural elements. Molecules consisting of one (a), two (b), or three (c) stem-loop elements. (d) A pseudoknot configuration.

whereas messenger RNA (mRNA) is shown as a linear thread lining up to be decoded. In fact, mRNA is known to fold into complex structures, and it is only "ironed out" by the passing ribosome assembly during translation. Recent advances in X-ray crystallography and nuclear magnetic resonance spectroscopy have opened up the investigation of RNA structure, and the number of resolved structures is increasing rapidly. Structural and sequence databases are beginning to reveal common motifs in RNA. Most such motifs are formed by conventional A-U, G-C base pairings, although occasional non-Watson-Crick pairings, such as G-G and G-A, can form by using alternate hydrogen-bond-forming sites. (For Activities 1 and 2, pairings other than A-U and G-C are and should be ignored.)

RNA's ability to fold in complex ways causes it to resemble proteins by having secondary and tertiary levels of structure. The determinants for folding these two molecules are quite different, however. Most proteins inside the cell are stabilized in a folded state with hydrophobic amino acid residues sequestered on the inside of the molecule, away from water, and by hydrophilic residues on the outside surface, exposed to the aqueous environment. Formal ionic interactions between amino acid side-chains of opposite charge also contribute to protein stability. Like RNA, proteins have many hydrogen bond interactions that help determine their shape. However, amino acids are not complementary in the manner of nucleotide bases, and thus there are no one-to-one pairing rules for amino acids. This fact makes it difficult to predict protein shapes from amino acid sequence alone. In the case of RNA, the accelerating accumulation of new sequence and crystallographic data makes the prospect look brighter for eventually predicting RNA three-dimensional shape from its sequence alone.

The folded structure of RNA is stabilized primarily by helical regions that form within the molecule based on Watson-Crick base pairing. Such helical regions differ in their degree of thermodynamic stability, depending on the number and nature of base pairs engaged in hydrogen bonding. Base stacking (the interaction between the electron clouds of the planar, cyclic bases when positioned on top of each other like a stack of dinner plates) also contributes to the stability of folded nucleic acids. Other more recently recognized contributors to folding and shape stability include the so-called ribose zipper, a juxtaposition of the minor grooves of two helical regions that is stabilized by hydrogen bonding of the 2′ OH group of ribose; the use of single-stranded loop regions that pair with "receptor" sequences elsewhere in the molecule; and the use of ions such as Mg^{++} to shield the uniform negative charge of the phosphate backbone. We next briefly describe some of the important structural motifs currently known to occur in folded RNAs, with emphasis on their known or proposed functions.

DOUBLE-STRANDED HELICES

With more than 50 percent of its bases in double-stranded form, a typical RNA contains a great deal of secondary structure. As mentioned previously, RNA helices are in the A-form whereas those in DNA are B-form. This difference in helix geometry creates significant differences in surface geography between RNA and DNA. Specifically, the major groove in RNA is quite deep and narrow, and the minor groove is shallow and wide, just the reverse of the B-form DNA helix. This difference in surface topography, along with the specific base sequence of the helix, determines the recognition and binding of other molecules to helical nucleic acids. For example, the shallow, wide, minor groove of RNA appears to be more accessible to protein side-chains and to present more hydrogen bonding opportunities than the major groove. Despite the different shapes of the two grooves, examples of protein binding to RNA appear to involve both to different extent. Because the binding of proteins to helical RNA and DNA can induce local bending and "melting" of base pairs, the limitations of helix groove size can be overcome to some extent. Indeed, there is some evidence that the single-stranded regions of the RNA, instead of the helices, may be more important as actual contact and recognition sites; the helical regions in this case serve to properly orient the single-stranded regions for presentation. Transfer RNA's single-stranded anticodon loop, which recognizes both mRNA and the synthetase enzyme that aminoacylates the tRNA, is a notable example of single-strand recognition ability.

HAIRPIN LOOPS

Like the anticodon loop, many important single-stranded recognition regions in RNA arise as part of a structure called a *hairpin loop*. The loop is created when a single strand of RNA bends back on itself to

form a double-stranded region. This creates a double-stranded stem and a single-stranded loop that caps the helix (see Figure 8). The number and size of hairpin loops vary among different RNA types; for example, the three loops in tRNA have 7–8 bases each, whereas the *Tetrahymena* self-splicing group I intron has six larger loops. So-called tetraloops, which have four unpaired bases atop their helical stem, are common in ribosomal RNAs (rRNAs); one such tetraloop in the 23S rRNA molecule appears to be the ribosome binding site of the toxic proteins ricin and sarin. In the *Tetrahymena* self-splicing intron, a tetraloop, along with its conserved receptor site elsewhere in the molecule, facilitates self-folding of the intron into the proper three-dimensional shape (Cate et al. 1996). Stem-loop structures are also found within the catalytic site of most ribozymes; for example, plant viruslike agents known as viroids contain self-cleaving RNA genomes whose catalytic site adopts the so-called hammerhead structure made up of three stem-loops. Smaller loops, known as *bulges*, are formed within a helical stem rather than at its end; they result when opposed bases are mispaired, causing them to pucker out from the helix.

There are many examples in which different aspects of gene expression, from mRNA transcription and translation to mRNA degradation, employ hairpin loops as control elements. Single-stranded loops may in general be preferred sites for the interaction of RNAs with regulatory molecules. For example, RNA is generally less susceptible to degradation by RNases when it contains a high proportion of hairpin secondary structure, and this can affect RNA's half-life in the cell. A well-studied example of structure controlling RNA degradation is the transferrin receptor mRNA. This mRNA encodes the cell surface receptor responsible for binding the plasma iron transport molecule, transferrin. A feedback mechanism responsive to plasma iron level increases the density of these transferrin receptors on the cell surface when iron levels are low, thereby enabling the cell to more efficiently scavenge transferrin-iron complexes from plasma. Operation of this system at the RNA level involves a regulatory protein, produced when iron levels are low, that binds to the transferrin receptor mRNA at a stem-loop structure near its 3′-end. The bound regulatory protein stabilizes the stem-loop, making the RNA less susceptible to degradation and thus able to be reused to make more copies of the receptor protein.

BASE TRIPLES

Unlike the two-way interaction within a standard pair of bases, a base triple is an interaction between three bases. Base triples are formed when a single-stranded region of an RNA nestles into the major or minor groove of a double-helical segment of the molecule; hydrogen-bonded triplets such as G-C-A can result. These triple-stranded regions help stabilize tertiary, three-dimensional structure and may be essential for certain RNA functions. For example, they have been found at the proposed catalytic regions in ribozymes such as the group I intron of *Tetrahymena*.

PSEUDOKNOTS

Pseudoknots represent a higher order, tertiary level of structure found in RNA. A pseudoknot results when some of the bases in an otherwise single-stranded loop pair with bases located outside that loop. This kind of interaction can potentially form a variety of distinct topologies, but all pseudoknots have two loops and two helical stems, usually with the stems sharing a common axis (see Figure 8). The multiple stems and loops of this RNA conformation provide more complex sites for interaction with proteins. In the 20 years since they were first identified, pseudoknots have been implicated in several examples of the regulation of gene expression. For example, pseudoknot structures in certain mRNAs of retroviruses, bacteria, and yeast appear to stimulate a gene-regulatory phenomenon called *ribosomal frame shifting*. In this process, pseudoknots in the mRNA, along with other specific base sequences, cause the ribosome to stall and slip during translation. The result is a change of reading frame that allows more than one protein sequence to be synthesized from a given mRNA. This is a clear example of an RNA structure that increases the compactness and efficiency of genetic information storage. Another gene regulation example, in bacteriophage T4, involves the binding of a transcription regulating protein to a pseudoknot in the promotor region of the gene 32 mRNA.

Pseudoknot structures also are found at the 3′-end of genomic RNAs of certain bacterial and plant viruses. Most intriguingly, these pseudoknots resemble the stem-and-loop configuration of tRNAs.

The exact function of these tRNA-like ends is unsettled, but they appear to be required for replicating of the virus RNA genome. Similar structures are found in a variety of other RNAs, such as the short molecules used to prime the reverse transcription of retroviral RNA genomes to cDNA copies, and the RNA transcripts made from certain fungal plasmids. Also, the chromosome-capping telomeres at the ends of eukaryotic chromosomes contain a TGG-rich sequence that is potentially able to base pair with the CCA sequence at the ends of these tRNA-like molecules. Weiner and Maizels (1987) have proposed the so-called genomic tag hypothesis, which posits that the tRNA-like sequence evolved early and functioned as a recognition tag that identified certain RNAs as genomes and somehow facilitated their replication. According to this view, the RNAs of similar shape found in today's viruses and cells can be considered molecular fossils—vestiges of a much earlier, RNA-dominated world.

These examples of RNA structure illustrate the general structure-function principle stated earlier: Polymer shape is determined by the linear order of subunits and their interactions within the molecule; three-dimensional shapes essential for the molecule's biological function emerge as a result. The structural motifs discussed in this section give only a hint of RNA's potential for shape diversity. Ongoing structural research employing a range of modern molecular techniques, such as RNA base sequence determination, site-directed mutagenesis, comparative analysis of sequences from different organisms, and advanced methods of X-ray crystallography, will no doubt reveal even more variety. That prospect is made all the more likely by the new RNAs that are being discovered in unexpected locations—and carrying out surprising functions—in the cell. In the next section, we discuss some of these newly recognized RNA functions.

The Diversity of RNA Function

Until recently, most biologists were aware of three types of cellular RNA: the standard trinity of mRNA, tRNA, and rRNA. Research largely within the last 20 years makes it clear that many more RNAs exist. They are found in diverse cellular locations, from the nucleus outward, carrying out a variety of key functions related to gene expression and metabolism. Some of the newly recognized RNAs are highly conserved between species as different as yeast and humans. And some, such as those bearing tRNA-like regions, appear to be related to familiar RNAs. The question of how these diverse and widely distributed RNAs arose and became so thoroughly integrated into cellular economy is central to the RNA world hypothesis.

Everything we know about evolution suggests that it is a conservative process that builds upon existing information to create novel structures and functions. That being so, if the RNA world hypothesis is correct and RNA was indeed the first information-encoding and catalytic entity, then we can predict that vestiges of those early RNA structures and functions, molecular fossils, should be present in contemporary organisms. Indeed, the diversity of RNAs and ribonucleotides in contemporary cells and their widespread involvement in key cellular functions provides much support for this hypothesis (Figure 9).

Most of the newly recognized RNAs share a common functional theme: gene expression. Unexpectedly, the roles of these molecules extend to aspects of gene regulation beyond simple coding and template functions. They play key roles at several levels, from participation in DNA replication and maintenance to regulation of transcription, RNA processing and editing, translation of mRNAs, protein localization within the cell, and modification of protein function. Because some of these RNAs are known to be catalysts, we next highlight some ribozymes and their functions.

RIBOZYMES

Seven categories of naturally occurring ribozymes have been found in the 15 years since their discovery (see Figure 5). They occur in a wide range of organisms, including viruses, bacteria, fungi, and plants; however, the prevalence of ribozymes in today's biological world is presently unknown. The relatively few catalytic roles apparently still left for RNAs may represent evolutionary vestiges, former roles having been taken over through natural selection by the later appearing but structurally more versatile proteins. Indeed, if RNAs were the first biological catalysts, as proposed by the RNA world hypothesis, they were far from the most efficient: Proteins catalyze thousands to millions of times faster. The kinetic data of ribozymes reflects this sluggishness; ribozymes have quite low Km values,

Figure 9 Lines of evidence supporting the RNA world hypothesis.

- RNA is informational *and* catalytic *in vivo*; no other biomolecule has both properties.
- The nucleotide sequences of RNAs common to all organisms (for example, rRNAs) are highly conserved (similar) among the many different species studied, suggesting that RNA was a key molecule present early in evolution.
- RNA or ribonucleotides are involved in most critical cellular functions in all three domains of life:
  - Adenosine triphosphate (ATP) is a universal energy carrier.
  - Universal metabolic pathways employ adenine nucleotide coenzymes (NADH, NADPH, FAD, CoA).
  - Protein synthesis employs mRNAs, rRNAs, and tRNAs.
  - rRNA by itself can catalyze peptide bond formation.
  - DNA synthesis requires the prior conversion of ribonucleotides to their deoxy form.
  - The ribonucleotide uracil, found only in RNA, is the precursor for DNA's thymine.
  - RNA is the primer for DNA replication.
  - Ribonucleotide derivatives function as key signaling molecules in the cell (for example, cAMP, ATP).
- RNAs function as primers in DNA replication and in reverse transcription of retroviral genomes.
- tRNA-like molecules are involved in nontranslational (nonprogrammed) polymerizations (for example, cell wall synthesis, polypeptide antibiotic synthesis).
- A tRNA-like molecule may have given rise to the RNA component of telomerase, the enzyme that maintains the ends of chromosomes.
- Enzymatic processing of mRNAs involves other small RNAs (snRNPs, RNase P).
- Protein sorting into the endoplasmic reticulum of all eukaryotes involves RNA (SRP-RNA).
- Ribonucleotides are used to activate and carry sugars during polysaccharide synthesis.

indicating very high affinity for their substrates and easy saturation. These kinetics seem well suited to ribozyme function, however; unlike protein enzymes, they typically catalyze only one reaction cycle (for example, their own removal from a larger molecule). Only two catalytic reaction mechanisms are known for naturally occurring ribozymes, transesterification and hydrolysis, and both employ OH groups as nucleophiles for cleavage of the RNA phospohodiester backbone. Notable examples are the coupled endonuclease-ligase reactions involved in splicing, and the coupled endonuclease-phosphatase phosphotransfer reactions that remove a substrate 3′ P and transfer it to the ribozyme.

Among the best-studied catalytic RNAs are the group I and group II introns, which autocatalyze their own excision from a larger precursor RNA and ligate the flanking exons. These introns are encoded in the genomes of a wide assortment of organisms, including bacteria, fungi, and plants. In eukaryotes, they are more commonly found in organelle genomes, such as fungal and plant mitochondria, and plant chloroplasts.

Group I and group II introns are rich in stem-loop secondary structure, particularly in their catalytic regions, and their higher order tertiary shape (at least in the case of the group I introns) appears to be stabilized by divalent cations like Mg^{++}. Most group I introns do not appear to require the aid (catalytic or otherwise) of proteins to self-splice. In contrast, the splicing of group II introns appears to benefit, both *in vitro* and *in vivo*, from the aid of maturase proteins that presumably help stabilize the correct tertiary structure of the intron required for self-splicing. Interestingly, at least some of these maturase proteins are encoded within the sequence of the intron itself; the intron carries coding "software" needed to make "hardware" that facilitates its catalytic role. Likewise, certain mobile introns capable of transposing to new genomic locations encode proteins, such as endonuclease and reverse transcriptase, that facilitate their movement. Enzymatic and coding abilities such as these are a far cry from the view held until quite recently that introns are nonfunctional genetic baggage or "junk DNA."

Unlike the autocatalytic processing of group I and group II introns, the processing of most eukaryotic mRNA precursors does not involve self-splicing introns. These introns instead require for their removal the actions of complex RNA-protein assemblies in the nucleus, known as spliceosomes. Spliceosomes are in effect macromolecular "splicing machines" and are reminiscent of ribosomes in being RNA-protein assemblies. The relative roles of the RNAs and proteins of spliceosomes, that is, which are catalysts and which are the structural elements, are not yet known. However, the catalytic

mechanism of splicing by spliceosomes resembles that of a group II intron, leading to speculation that the several spliceosomal RNAs may be an "intron-in-pieces": fragmented descendants of a once self-splicing intron.

RNA EDITING

Another level of gene expression regulation that involves the functional RNAs is the process of RNA editing. During editing, particular bases in the precursor RNA are added, deleted, or changed following transcription. The resulting edited RNA has a base sequence different from that encoded in the DNA and transcribed RNA precursor. Most examples of RNA editing involve eukaryotic mRNA precursors, and when these are edited the mature mRNA specifies a different protein product than that encoded in the gene. RNA editing thus increases the number and diversity of products from a given gene. As with gene regulatory processes like alternative splicing and *trans*-splicing, RNA editing effectively increases the information coding capacity of a genome. Although the detailed chemical and catalytic mechanisms of editing are not yet known, one version of the process employs small RNAs known as *guide RNAs*. The base sequences of the guide RNAs enable them to pair with the pre-edited target RNA and specify the bases to be added, deleted, or altered by complementary pairing.

Two additional examples illustrate the variety of RNA effects on gene expression. During the translation process of protein synthesis, a peptidyl transferase catalytic activity covalently links incoming amino acids to the growing polypeptide. This catalytic activity is known to reside in the large ribosomal subunit, which in *E. coli* consists of some 31 proteins and 2 distinct RNA molecules. Which of these ribosome components catalyzes peptide bond formation has long been a mystery. The conventional assumption was that it must be one of the proteins. Recent results, however, strongly suggest that the 23S rRNA component is the catalyst. Noller et al. (1992) extracted more than 95 percent of the proteins from the large subunit, leaving the 23S rRNA, and the peptidyl transferase activity, intact; treatments that damage the RNA eliminate the activity. More recently, Nitta et al. (1998) demonstrated that cloned segments of the 23S rRNA never exposed to ribosomal proteins could be reconstituted in the test tube and could catalyze peptide bond formation. The emerging picture is that of an RNA catalyst in the lead role of a reaction essential for life, with ribosomal proteins playing structural supporting roles. This view is consistent with the notion that complex ribosomes evolved when functions once carried out solely by RNAs were improved upon by the addition of proteins.

RNAs also play a role in the cellular localization of newly made proteins. In eukaryotes, a complex translocation machinery enables nascent polypeptides bearing an amino-terminal "signal sequence" to cross into the lumen of the endoplasmic reticulum (ER). A key element of this translocation machinery is the so-called signal recognition particle (SRP), a cytoplasmic RNA-protein assembly consisting of a core 7S RNA, to which six different protein components bind. The signal recognition particle binds to the free-floating ribosome-nascent peptide complex. It then carries the complex to the surface of the ER membrane, where the SRP binds to a receptor. The SRP in effect acts as a bridge to tether the ribosome-peptide to the surface of the ER, and thus facilitates movement of the newly made polypeptide into the lumen. The 7S RNA is necessary for SRP function and may be the component that binds directly to the ER membrane receptor protein.

RNA and Evolution: Molecular Evolution in a Test Tube

A basic assumption of the RNA world hypothesis is that early in the evolution of life, one or a few molecules came to dominate the pool of RNAs that had been randomly generated by nonbiological processes. Assuming that RNA's appearance preceded that of catalytic proteins, the RNA must have been able to replicate itself in order for evolution at the molecular level to get under way. The demonstrated catalytic ability of RNA makes the possibility of self-replication much more plausible. Even though no RNA capable of catalyzing its own replication has yet been found in nature, recent experiments in directed evolution (molecular evolution in a test tube) have demonstrated that RNA does indeed have this potential.

Sol Spiegelman and his group first demonstrated evolution and selection of RNA molecules in a test tube in the 1960s. Spiegelman's *in vitro* experiments

started with a mixed population of RNA genomes from the bacteriophage Qβ, which were transferred serially to a series of tubes supplying only ribonucleotides and replicase (the RNA polymerase that copies the viruses' RNA genome inside host cells). By limiting the time available for replication in each tube, the experiments imposed speed of RNA replication as the selection criterion. Several cycles of *in vitro* replication and transfer led to variant RNAs that could be copied at greatly increased speed. By altering the physical or chemical selection conditions, populations of RNA adapted to the imposed conditions came to predominate in the mixture.

The modern era of *in vitro selection* (also referred to as directed evolution) began in 1989–90 and was made possible by technical advances in RNA and DNA synthesis (for example, automated oligonucleotide synthesis), nucleic acid amplification (for example, polymerase chain reaction), and selection methods (for example, affinity chromatography). Starting with a large synthetic population (pool) of randomly varying RNA or DNA molecules, the goal of *in vitro* selection experiments is to amplify those variants that are able to meet some experimentally imposed selection criterion, such as the ability to bind to a particular target molecule or catalyze a particular chemical reaction. After 10 to12 cycles of selection and amplification with mutation, molecules well adapted to the selection criterion predominate in the pool (Figure 10).

A parallel can be drawn between the variation, selection, and amplification aspects of these *in vitro* experiments and natural, biological evolution. However, the action of selection in these experiments is more direct than in biological evolution: In organisms, molecular structure-function is selected somewhat indirectly through a complex organismal phenotype; with *in vitro* methods the molecule's base sequence, which directly determines its function, is the selected phenotype. It might be said that the generation of highly functional molecules from a random assemblage of sequences is akin to a tornado assembling a 747 from random parts, creationist views notwithstanding.

The RNA and DNA molecules found by this powerful experimental approach tell us a lot about the potential range of nucleic acid functions and their evolutionary potential. In fewer than 10 years, the method has uncovered many new synthetic ribozymes, supplementing the seven known natural classes. Significantly, one of these synthetic RNAs can copy an RNA template, forming short complementary strands. Further laboratory refinements of this molecule may yield a bona fide RNA-dependent RNA polymerase, a key ribozyme activity required for self-replication and assumed in RNA world ideas about the origin of life.

In vitro evolution experiments selecting for *DNA* molecules have shown that single-stranded DNA, like RNA, has catalytic potential. Perhaps the distinction

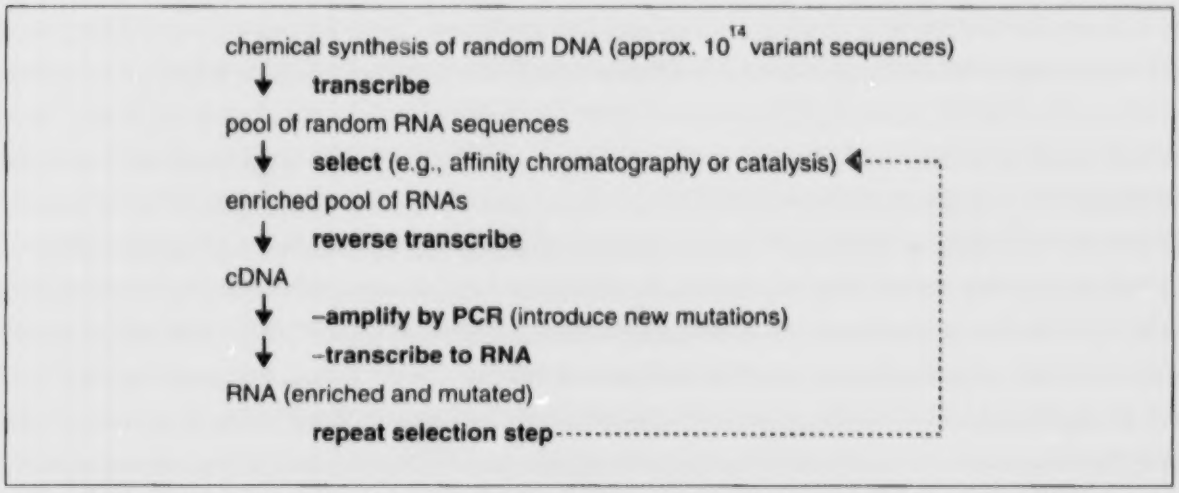

chemical synthesis of random DNA (approx. 10^{14} variant sequences)
transcribe
pool of random RNA sequences
select (e.g., affinity chromatography or catalysis)
enriched pool of RNAs
reverse transcribe
cDNA
–amplify by PCR (introduce new mutations)
–transcribe to RNA
RNA (enriched and mutated)
repeat selection step

Figure 10 The strategy of directed evolution *in vitro*. Starting with a large, randomly synthesized population of nucleic acid, repeated cycles of selection followed by amplification can yield a particular sequence well suited to the selection criterion.

between RNA and DNA is not as great as we have come to believe. Recall that the only chemical difference is one extra oxygen atom in RNA (in the 2' OH group of ribose). Also, both RNA and single-stranded DNA can fold into complex shapes. If there really is little difference between these two closely related molecules, why is DNA today the storage form of most genetic information, and why is RNA the active form performing a variety of cellular tasks? The current thinking is that DNA appeared later in the RNA world, most likely as an RNA derivative, and DNA's succession to the role of information repository may be related to its greater chemical stability (due to its fully double-stranded configuration and lack of a 2' OH nucleophile). Or perhaps DNA proved to have too small a range of catalytic ability to out-compete RNA in a functional role. It remains an intriguing question whether life could have evolved equivalently if the earliest catalytic events had been DNA-based and RNA had appeared later as the storage form.

RNA, Health, and Disease

Many viruses have RNA genomes. Some are important human pathogens: polio, HIV, flu, and measles, among others. Numerous examples are also found among plant viruses, several of which have major economic impacts in agriculture and forestry. A common feature of all RNA viruses is a high level of variation among their genome sequence, the result of a high mutation rate (10^{-3}–10^{-4} per base pair per replication). The frequency with which mutations arise in a population is determined in part by how often the organism reproduces. Viruses replicate millions of times each day, so random mutations are constantly arising. Another factor affecting the frequency of mutation is the fidelity of the replication process itself. The replicase enzyme that copies RNA genomes occasionally makes random errors, inserting the incorrect monomer (for example, A opposite G, or U opposite C). Indeed, the polymerases that copy DNA also make errors, but cells have evolved a molecular quality-control, proofreading mechanism that can correct mistakes most of the time, keeping the DNA mutation rate low (less than 10^{-7}). RNA lacks such a correction mechanism, a condition that may be related to its more limited role as a genome in nature. Although populations of RNA viruses have greater variation in genome sequence than DNA viruses, variation is essential to the adaptation and evolution of all viruses, indeed, all life forms.

Because of this high mutation rate, the many progeny viruses made in an infected cell constitute a population of variants, with typically one to three changes in each viral genome. The genome sequence of a population of RNA viruses thus is not unique, but rather is a population of variant "quasispecies." Another source of genetic variation in viruses is their ability to exchange whole blocks of genes among progeny by the process of recombination. Most of the genome alterations are deleterious, and those viruses will not survive. Other genome changes, however, can confer advantages for different aspects of viral behavior: faster replication, increased virulence, or decreased sensitivity to antiviral drugs.

In the case of HIV, it has been estimated that each of its 10,000 RNA bases is mutated more than 10,000 times each day in an infected person (Coffin, 1995). HIV's particularly rapid mutation rate and the chronic nature of the infection make it possible to detect the emergence within an individual patient of new mutant strains during the course of the disease. The virus thus can evolve within the microcosm of a single human host. In similar ways, new strains of cold and flu viruses continue to emerge and plague humankind.

With both RNA and DNA viruses, the war between the hosts and the viruses that infect them is sophisticated, subtle, and evolving: The host inactivates the virus or limits its propagation, the virus subverts the cell to make more virus. The battle shifts back and forth with move and countermove. A critical move by the host is to mount an immune response against the virus. Unfortunately, the RNA virus's ability to rapidly generate genetic diversity produces an ever-moving target for the immune system, as well as for developers of vaccines and antiviral drugs. In the case of HIV, the virus has evolved an additional powerful strategy to ensure its persistence in the host, that of attacking and disabling the immune system itself.

Some viruses subvert cellular defenses by employing RNA molecules as weapons. The cellular interferon response is one example. When infected with virus, cells typically react by secreting the potent signaling molecule interferon. This protein protects neighboring cells from becoming victims by temporarily shutting down their protein-synthetic machinery. This clever cellular defense strategy

denies the virus access to the one source of new proteins needed to complete its life cycle. Adenovirus, however, has evolved a more clever way to undermine this defense: Specialized viral RNA molecules (for example, VAI RNA) effectively block interferon's action and prevent the shutdown of cellular protein synthesis. The cell has no option but to contribute slavishly to the production of viral progeny.

Viruses can likewise defeat medical therapies. Although vaccines have been dramatically successful in curtailing some viral diseases, such as polio, they are less effective at combating the rapidly evolving RNA viruses, such as rhinovirus (common cold), influenza, and HIV. For these, the current approach is to develop antiviral drugs, chemical agents that either block entry of the virus into host cells or, once inside, block key steps in the viral reproduction cycle. Current treatment of HIV, for example, employs drugs such as Acyclovir, AZT, and protease inhibitors. By targeting different aspects of viral reproduction, these drugs are proving effective at slowing the course of HIV infection.

Evolutionary considerations, however, dictate that the *way* in which these drugs are administered is a key factor in their long-term effectiveness. Administered singly, any given drug soon loses effectiveness as the rapidly mutating virus generates a population that is resistant. If the drug is replaced by another in an attempt to hit the remaining resistant population, that population will in turn give rise to a new population that is resistant to both drugs . . . and so it goes. The unintended but predictable outcome of this type of serial drug treatment is the eventual creation of a viral population that is resistant to several drugs. Such a multiply resistant virus could spread quickly in the host population. A much better strategy, one that makes rational use of evolutionary principles, is to use all three drugs in combination. The key to this approach is that each drug *independently* and *simultaneously* hits a different step in the virus's reproduction cycle. For a triply resistant viral population to emerge in the presence of all three drugs, three independent resistance mutations would have to occur together in a single founder *virus*, a much less probable occurrence.

The alternative to devising separate drug treatments for each different type of virus is to design a broad-spectrum antiviral agent that is effective against many viruses. Unfortunately, because viruses are dependent on host cell machinery (such as that for protein synthesis), any drug that interferes with a key synthetic step common to all viruses likely also would inhibit cellular function. The situation is different for bacteria, which have evolved enough differences at the molecular level from eukaryotic cells, including RNAs, that it is possible to develop drugs that specifically target bacteria. Agents such as penicillin, which impairs synthesis of the unique molecular outer layers of bacterial cells, were dramatically effective when first introduced. Also, the protein synthetic apparatus of bacteria is sufficiently different in molecular detail from that of eukaryotes that it is a good target for antibiotics. Some of these antibiotics inhibit bacterial protein synthesis by binding directly to RNA components of the bacterial ribosome. Structural differences between bacterial and nonbacterial rRNAs account for the specificity of the antibiotic in this case. Mutations that change the structure of the bacterial rRNAs, such that the antibiotics no longer bind, confer resistance on the pathogen. Once again, we see that the ongoing process of evolution, fueled by mutations such as these in the bacterial and viral populations, ultimately undermine the effectiveness of even the most powerful therapeutic agents in our arsenal. Activity 4 of the module explores RNA's role as a target for certain antibiotics and antiviral agents, as well as the acquisition of resistance.

Two novel strategies focusing on RNA with potential application against viral pathogens are currently being explored in the laboratory: *in vitro selection* and *antisense technology*. The former approach employs the *in vitro* selection strategy described above to select functional nucleic acid molecules that bind to and inactivate viral components. Ribozymes able to cleave DNA have been developed, as have deoxyribozymes able to cleave RNA. Looking ahead, it may one day be possible to engineer into the genome of affected cells a ribozyme sequence that has been selected *in vitro* for its ability to inactivate the RNA, DNA, or protein of a viral pathogen. The coding sequence for a ribozyme having, for example, anti-HIV activity might be incorporated into a population of the patient's immune cells. Subsequent expression of the ribozyme "bullet" in these cells and their descendants conceivably could have significant therapeutic effects against the virus.

Another potential application of *in vitro*-selected nucleic acids is as ligand-binding molecules; such molecules could be useful as biochemical reagents or as potential therapeutic agents. The term *aptamer* describes *in vitro*-selected RNAs or DNAs that have the ability to bind with specificity to another molecule (aptamers, unlike ribozymes, are not catalytic). Aptamer sequences with unique shapes have been developed to bind a wide range of ligands, including small organic dyes, coenzymes, amino acids, vitamins, and viral proteins. The binding specificity of aptamer RNAs is high enough that they can distinguish between ligands as similar as theophylline and caffeine, which differ by only a single methyl group. These designer nucleic acids promise to be useful additions to the growing list of functional biomolecules, including protein enzymes and antibodies, that already have been generated by *in vitro* techniques.

The second RNA-related approach to therapeutics, *antisense technology*, attempts to inactivate and neutralize unwanted mRNAs that derive from a mutated target gene. This approach takes advantage of the tendency of single-stranded RNAs to bind to sequences complementary to themselves. The strategy is to engineer into affected cells an antisense copy of the defective gene, that is, one in which the 5'-3' orientation of the gene has been reversed relative to its promotor. Insertion of the inverted gene copy creates a situation in which RNA transcripts from the inverted gene, known as *antisense transcripts*, arise from what is normally its nontranscribed coding strand. The antisense transcripts are thus complementary to the sense transcripts from the defective gene, which arise from the opposite, template strand. Antisense and sense RNA molecules can bind and neutralize one another, effectively silencing the expression of the defective target gene, perhaps a provirus or an oncogene. Successful therapeutic applications of both the *in vitro* selection and antisense approaches await advances in gene therapy technologies needed to incorporate engineered sequences into the genome in a functional state.

References and Related Literature

Abbott, I., et al. (1977). Comparative ecology of Galápagos ground finches (*Geospiza* Gould): Evaluation of the importance of floristic diversity of interspecific competition. *Ecological Monographs, 47*: 151–184.

Cairns-Smith, A.G. (1985). *Seven clues to the origin of life*. New York: Cambridge University Press.

Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Szewczak, A.A., Kundrot, C.E., Cech, T.R., & Doudna, J.A. (1996). RNA tertiary structure mediation by adenosine platforms. *Science, 273*: 1696–1699.

Cech, T.R. (1998). Structure and mechanism of the large catalytic RNAs: Group I and group II introns and ribonuclease P. In Gesteland, R.F., & Atkins, J.F. (Eds.), *The RNA world* (pp. 239–269). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Cleland, A., Watson, H.G., Robertson, P., Ludlam, C.A., & Brown, A.J.L. (1996). Evolution of zidovudine resistance-associated genotypes in human immunodeficiency virus type 1-infected patients. *Journal of Acquired Immune Deficiency Syndromes and Human Retrovirology, 269*: 6–18.

Coffin, J.M. (1995). HIV population dynamics *in vivo*: Implications for genetic variation, pathogenesis, and therapy. *Science, 267*: 483–489.

Ekland, E.H., Szotak, J.W., & Bartel, D.P. (1995). Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science, 269*: 364–370.

Eldredge, N. (1995). *Reinventing Darwin: The great debate of the high table of evolutionary theory*. New York: John Wiley & Sons, Inc.

Ellington, A.D., & Joyce, G.F. (1990). *In vitro* selection of RNA molecules that fold into specific ligand-binding ligands. *Nature, 346*: 818–822.

Gesteland, R.F., & Atkins, J.F. (Eds.). (1998). *The RNA world* (2nd ed.). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Gilbert, W. (1986). The RNA world. *Nature, 319*: 618–620.

Howard Hughes Medical Institute Holiday Lecture on Science. (1995). *The Double Life of RNA* [videotape set]. (A four-tape series of lectures by Nobelist Thomas Cech, discussing his discovery of ribozymes and other important functions of RNA. Summaries of the lectures can be found online at http://www.hhmi.org/.)

Jenison, R.D., Gill, S.C., Pardi, A., & Polisky, B. (1994). High resolution molecular discrimination by RNA. *Science, 263*: 1425–1429.

Joyce, G.F. (1992, December). Directed molecular evolution. *Scientific American, 267*(6): 90–97.

Joyce, G.F., & Orgel, L.E. (1993). Prospects for understanding the origin of the RNA world. In Gesteland, R.F., & Atkins, J.F. (Eds.), *The RNA world* (pp. 1–25). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

BEST COPY AVAILABLE

Landweber, L.F., Simon, P.J., & Wagner, T.A. (1998). Ribozyme engineering and early evolution. *BioScience, 48*(2): 94–103.

Mills, D.R., Peterson, R.L., & Spiegelman, S. (1967). An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proceedings of the National Academy of Sciences, 58*(1): 217–224.

Nitta, I., Kamada, Y., Noda, H., Ueda, T., & Watanabe, K. (1998). Reconstitution of peptide bond formation with *Escherichia coli* 23S ribosomal RNA domains. *Science, 281*: 666–669.

Noller, H.F., Hoffarth, V., & Zimniak, L. (1992). Unusual resistance of peptidyl transferase to protein extraction methods. *Science, 256*: 1416–1419.

Orgel, L.E. (1994, October). The origin of life on the earth. *Scientific American, 271*(4): 77–83.

Simons, R.W., & Grunberg-Manago, M. (Eds.). (1998). *RNA structure and function*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Smith, J.M., & Szathmary, E. (1995). *The major transitions in evolution*. Oxford: W.H. Freeman (Spektrum).

Szostak, J.W., & Ellington, A.D. (1993). *In vitro* selection of functional RNA sequences. In Gesteland, R.F., & Atkins, J.F. (Eds.), *The RNA world* (pp. 511–533). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Unrau, P.J., & Bartel, D.P. (1998). RNA-catalyzed nucleotide synthesis. *Nature, 395*: 260–263.

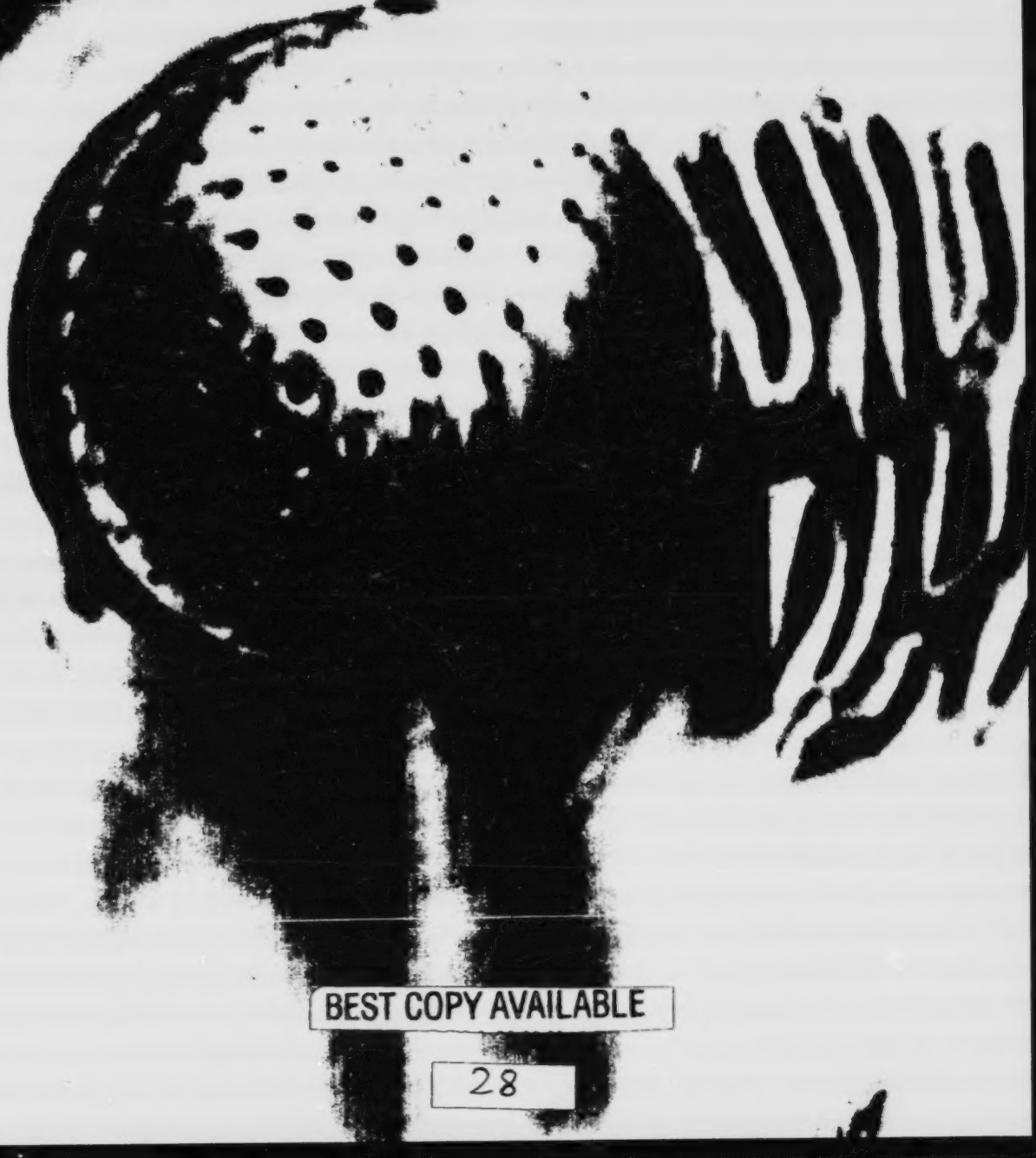
Weiner, A.M., & Maizels, N. (1987). 3′ Terminal tRNA-like structures tag genomic RNA molecules for replication: Implications for the origin of protein synthesis. *Proceedings of the National Academy of Sciences, 84*: 7383–7387.

Wright, M.C., & Joyce, G.F. (1997). Continuous *in vitro* evolution of catalytic function. *Science, 276*: 614–617.

Wyatt, J.R., & Tinoco, I. (1993). RNA structural elements and RNA function. In Gesteland, R.F., & Atkins, J.F. (Eds.), *The RNA world* (pp. 465–496). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Zaug, A.J., McEvoy M.M., & Cech, T.R. (1993). Self-splicing group I intron from *Anabaena* pre-tRNA: Requirements for base-pairing of the exons in the anticodon stem. *Biochemistry, 32*: 7946–7953.

Annotated Faculty Pages



BEST COPY AVAILABLE

About the Activities

Faculty Introduction

RNA, like DNA and protein, is a polymer that carries out essential biological functions. As with DNA, RNA has long been appreciated as an informational molecule, storing or transferring information encoded in the sequence of its monomer nucleotide bases. Surprising recent discoveries show that many naturally occurring RNA molecules also mimic a well-known function of proteins: biochemical catalysis. This discovery greatly expands RNA's biological role from its traditional depiction in the central dogma of molecular genetics. RNA is a genetic molecule of action.

The variety of different RNA functions being revealed in cells raises interest in its three-dimensional structures. The authors of the article "RNA Structural Elements and RNA Function," in the book *The RNA World*, describe the significance of and search for the relationship between RNA structure and function:

> The structural elements that now exist in RNA must have evolved to provide chemical stability and to facilitate their biological functions. Structural biologists believe that determining the conformation and stability of the structures adopted by RNA will provide a better understanding of their functions. This belief is based mainly on the success of Watson and Crick with double strands for DNA . . . we continue to be optimistic and to study RNA structure in the hope that it will provide hints about how RNA does its many jobs. The main value of a structure is to suggest new experiments. The structure allows predictions about the effect of mutations, inhibitors, and enhancers, and about the mechanisms of the reactions. The overall goal is to learn the general features of RNA structure that can be inferred from the sequence and then to relate these structures to biological function.

Single-stranded RNA folds into elaborate shapes that are stabilized by base-pairing interactions similar to those that connect the strands of a DNA double helix. Regions of RNA that are far apart in the primary sequence can come into proximity and form hydrogen-bonded helical "stems" that bear single-stranded "loops" or other complex shapes. These loops can subsequently bond to other single-stranded regions of the RNA sequence until a three-dimensional, functional shape results. Not all of the three-dimensional structures a molecule can assume are necessarily functional; however, there is a clear relationship between a particular RNA molecule's

BEST COPY AVAILABLE

structure and its specific function. You will find additional information on RNA shape determination in the *Faculty Background*.

Strategies for Teaching the Activities

These activities present opportunities for students to reason and to use observational data. The activities combine a series of hands-on modeling exercises with accompanying Challenge Questions and brief For Your Information (FYI) essays. The introductions to the modeling exercises provide a minimum of background information, just enough for students to explore the concepts of RNA shape determination in an active, thought-provoking, *inquiry* mode. The FYI essay for a given part of an exercise typically follows that part and is meant to be read after the students have actively grappled with the concepts. The essay provides students with additional, clarifying information about the concepts. If class time is limited, you may assign the essays and certain questions as postexercise homework.

The *Student Pages* are masters that you can copy and distribute to students. The *Annotated Faculty Pages* contain the student text (in **bold type**) and hints for teaching the activities, answers to Challenge Questions, optional extension exercises, and reagent preparation instructions (in regular type). The activities are designed for groups of four students, but can be accomplished successfully by individual students. The *Copymaster* and *Template* sections contain masters needed for the exercises.

Activity 1
RNA Structure: Tapes to Shapes

Activity at a Glance

Exercise 1.1:	Students compare primary base sequences and search for similarities (homologies). Students attempt to envision higher order structure without the aid of models.
Exercise 1.2:	Students use models and folding rules to see the relationship between the primary sequence of monomers and the three-dimensional structure of the RNA polymer.
Exercise 1.3a:	Students explore the relationship between RNA structure and function by applying a functional test to their models.
Exercise 1.3b:	Students model the effects of mutation by changing the primary sequence and determining the effects on structure and function.
Exercise 1.4:	Students use a computer program to confirm their RNA structures.
Extension Exercises:	Students participate in paper and pencil activities that explore the versatility of long RNA molecules.

Concepts

- The linear sequence monomers determine a polymer's three-dimensional shape.
- The biological function of polymers such as RNA requires the appropriate three-dimensional shape.
- Scientists use modeling for discovery as well as for demonstration.
- Mutation produces heritable changes in RNA structure and function.
- Molecular function is subject to natural selection.

BEST COPY AVAILABLE

Focus

Students are challenged to make the connection between molecular structure and biological function. To do this, students model RNA structure (linear, secondary, and tertiary) using pipe cleaners, a set of base-pairing and folding rules, and RNA primary sequence data. The activity introduces the idea that molecular function can be subject to natural selection (a concept explored in depth in Activity 3).

Connections to Lecture Topics

Use Activity 1 to introduce molecular biology or RNA, or with lectures on molecular structure-function relationships or mutation.

Estimated Time

100 minutes (depends on whether students color-code models in class or start with color-coded models)

Materials Preparation

For each team of four students, provide

- 16 white pipe cleaners (also called *chenille stems*), 30.5 cm long. An alternative model is described in Exercise 1.2 that uses clear, unlabeled vinyl tubing (i.d. 1/4 inch, o.d. 3/8 inch). If you choose this model, each student will need 2 50-cm lengths of tubing, 10 twist-ties, and 2 preprinted strips of paper (see *Copymasters*).
- yellow, green, and black markers, 1 per student. Use to color-code the models; fewer markers will work if students within a group share, but it may slow them down.
- template sheets for color-coding (see *Templates*)
- clear, dull tape
- 4 premarked styrene balls (5 cm in diameter). Directions for marking the balls are given under Exercise 1.3a.
- 1 metric ruler per student
- access to the World Wide Web

ANNOTATED STUDENT ACTIVITY

Introduction

What does the shape of a thing have to do with the job it performs? You no doubt can think of a variety of familiar household objects that clearly show the relationship between shape and function, such as a metric wrench that fits only metric-sized bolts or a key that must have exactly the right shape to unlock a lock. Molecules also display an essential relationship between structure and function.

This activity offers you a combination of hands-on exercises and short essays that examine structure and function in RNA. To discover for yourself the answers to the Challenge Questions posed in an exercise, you should perform the hands-on inquiry portions of the exercise first. Then read the accompanying For Your Information (FYI) essays to help you understand your findings. In Activity 1, you will construct models of different RNA molecules to investigate the following:

- the principles underlying shape determination in RNA,
- the variety of shapes in RNA,
- how shape can influence function in RNA,
- how alterations (mutations) in RNA building blocks can influence structure and function, and
- how comparisons among related molecules from different species can be useful.

Exercise 1.1: RNA Sequences: Primary Structure

Procedure

Read the following text, then answer the Challenge Questions.

We can use the analogy of an elongated tape as a good starting point for exploring the complex shapes of RNA molecules. RNA molecules are relatively large *polymers*, molecules composed of many smaller *monomer* subunits; the tape is analogous to the *linear sequence* of building blocks that make up an RNA molecule. You may already be familiar with proteins as biological polymers; their primary structure consists of long chains of covalently linked amino acid monomers.

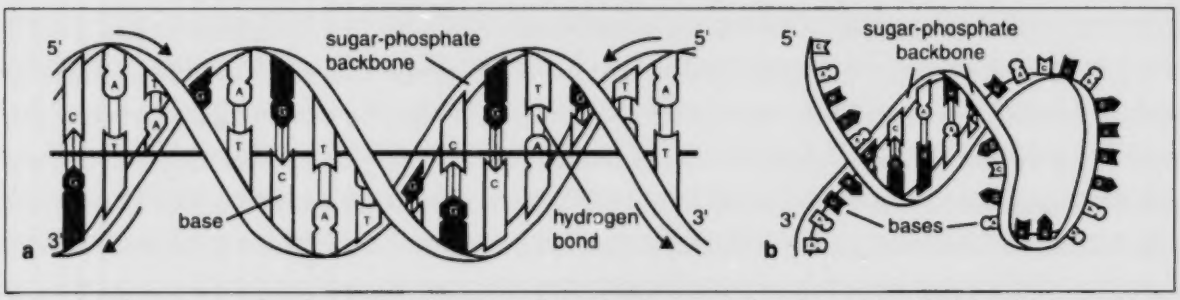


Figure 1.1 The structure of nucleic acids. (a) In DNA, alternating sugar and phosphate molecules form a linear backbone joining nucleotide subunits into chains. Two chains are held together by specific base pairing and coiled around a central axis to form a double helix; the strands run in opposite, antiparallel directions. (b) A single linear RNA strand folds by undergoing specific base-pairing interactions (A to U and G to C) within the strand.

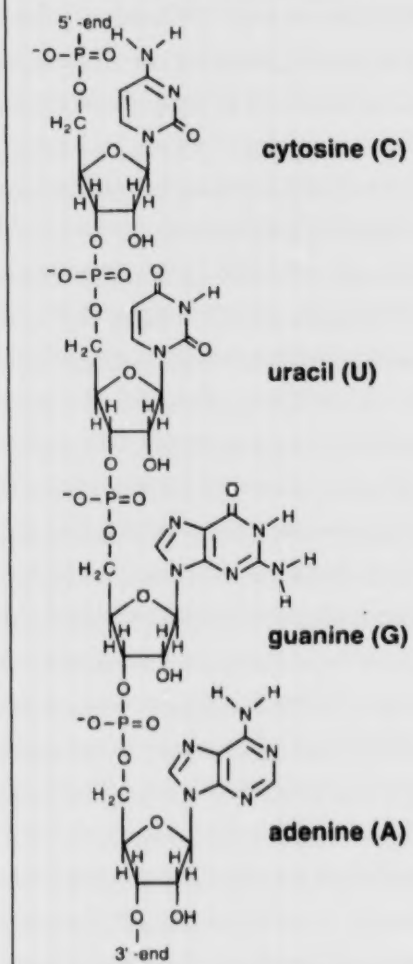


Figure 1.2 The primary sequence of a short section of single-stranded RNA illustrates the chemical structure and linkage of the monomer nucleotides.

In the case of nucleic acid polymers, RNA and DNA, the monomer subunits are called *nucleotides* and are linked together covalently into a long chain or tapelike structure (Figure 1.1). Different RNA molecules may be made up of tens to thousands of nucleotide monomers. The nucleotides themselves are made up of smaller components, namely a sugar molecule (ribose in RNA, deoxyribose in DNA), a phosphate, and a nitrogen-containing base. The base is the "business end" of the nucleotide; it interacts with other nucleotide bases in precise ways. Four different types of nucleotide occur in RNA, each having a different base: adenine (A), guanine (G), cytosine (C), and uracil (U). (In DNA, the base thymine (T) replaces uracil (U). Scientists typically refer to the individual nucleotide monomers of RNA and DNA simply as bases.

The linear order of bases in a nucleic acid is referred to as its *primary structure* (1° structure) (Figure 1.2). Recently developed biochemical methods for sequencing nucleic acids make it possible to determine in just hours the linear sequences of nucleotides in DNA or RNA that would have taken weeks to determine 15 years ago. This and several other technological breakthroughs have made possible the Human Genome Project, a worldwide effort to sequence completely the genetic material of *Homo sapiens*, as well as that of several other species. To date, the genomes of 25 bacterial species, yeast, the nematode worm *C. elegans*, and the fruit fly *Drosophila* have been sequenced completely.

Consider the nucleotide base sequence data for the four different RNA molecules shown in Figure 1.3.

Seq. 1	5′	U	A	G	A	G	U	C	U	G	A	U	C	G	A	C	A	G	G	U	G	C	U	C	U	A	C	C	G	A	A	C	C	U	G	3′
Seq. 2	5′	A	U	A	U	A	C	C	C	G	A	A	A	A	U	C	A	G	G	U	G	C	U	C	U	C	C	C	G	A	A	C	C	U	G	3′
Seq. 3	5′	C	G	U	C	G	G	A	A	A	C	G	U	A	U	C	C	U	U	U	A	C	A	C	C	G	A	A	G	U	A	U	A	C	G	3′
Seq. 4	5′	G	C	G	C	G	A	U	A	C	U	C	A	A	C	C	C	C	A	A	A	A	U	U	G	C	C	A	G	A	A	G	C	G	C	3′

Figure 1.3 Primary sequences of some RNA molecules.

BEST COPY AVAILABLE

Challenge Questions

1. **Do you see any similarities in primary sequence among any of the molecules in Figure 1.3? If so, indicate them by circling regions of sequence that appear to be conserved between molecules.**

   Starting at the 5′-end, all four sequences share bases at positions 15 (C), 28 (G), and 30 (A). In addition, sequences 1 and 2 have in common 19 of their 20 3′-most bases.

2. **At this point, what predictions can you make about the relative shapes of those molecules that have conserved regions of sequence?**

   No precise predictions can be made at this point. However, it is reasonable to conclude that sequences 1 and 2, which share a long stretch of bases, might be closer in shape to one another than to sequences 3 or 4.

3. **Can you mentally visualize the three-dimensional shape of each molecule?**

   At this point, prior to modeling the molecules, it is clearly not possible to mentally visualize the shapes of the molecules. It is for this reason that the modeling exercise that follows is of value.

4. **Do you think it is possible for two molecules with entirely *different sequences* to have the *same shape*? Discuss this within your team.**

   Although the answer to this question will likely not be obvious to the students, it *is* indeed possible for RNA molecules of different base sequences to assume the same shape.

The introductory text to Exercise 1.1 briefly reviews the concepts of *biological polymers* and the *linear sequence* orientation of their component subunits. Some terminology relating to polymers also is presented. Students learn the four *nucleotide bases* found in RNAs and are formally introduced to the concept of RNA *primary (1^0) structure*. Students go on to visually examine four printed RNA primary base sequences. Challenge Questions ask students to discern patterns of similarity among the four sequences. Students are then asked to try and envision what the three-dimensional shapes of the molecules might be. *Be aware that at this point in the exercise, students are working without the aid of molecular models and with no formal introduction to the rules that determine higher order shapes in RNA.* Let the students ponder the sequences just long enough to realize that they cannot answer the shape question without the aid of a model. Challenge Questions 3 and 4 are meant to be thought-provoking and to generate discussion within the group. Students should read the FYI essay *Modeling in Science after* completing this exercise and before starting Exercise 1.2.

FOR YOUR INFORMATION
Modeling in Science

In all likelihood, you were not able to envision the shape of the molecules in Figure 1.3 simply by studying their primary sequence. Large biological molecules often have complex shapes that are much too difficult to visualize from this kind of raw data or from a written formula. This is where modeling becomes necessary. Models are important tools in science, because they help us visualize structures we cannot see directly. Model building is particularly helpful in the study of biochemistry and molecular biology. You may have used molecular models in previous chemistry or biology classes to learn the structure of known molecules. Scientists also use models to discover *new* molecular shapes. Molecular models enabled Watson and Crick, for example, to discover and visualize the structure of DNA for the first time.

Figure 1.4 (a) James Watson and Francis Crick with their original wire and cardboard model of the DNA double helix. (b) RNA researcher Harry F. Noller. (c) Computer graphic models of two transfer RNA (tRNA) molecules.

A variety of modeling materials have been used over the years, from the simple wire and cardboard models used by Watson and Crick for their work with DNA, to sophisticated computer-generated images (Figure 1.4). Research scientist Harry F. Noller, at the University of California—Santa Cruz, used simple pipe cleaners to begin to model some important features of RNA structure.

Before a structural model of any molecule can be built, scientists must first gather a lot of basic data about its chemical composition and physical properties. Today new biochemical analyses, such as nucleotide sequencing, are combined with advanced methods of X-ray crystallography and spectroscopy for this purpose. Computer modeling of molecular structure enables scientists to rotate and observe virtual molecules in three dimensions, allowing rapid comparisons to known molecules. Such comparisons often provide new clues to the behavior of the molecule under study. Some modeling software allows the scientist to change atoms or molecular subunits in the model or change incubation conditions to determine the effects on the molecule's structure. Examples of computer-generated models are available on the World Wide Web. You might try the following sites:

- The RNA World (IMB-Jena) at http://www.imb-jena.de/RNA.html (an extensive site with many links to sequence and structural databases, modeling software, and books and tutorials)
- RNA Pages at http://www.kwl.t.u-tokyo.ac.jp/~ueda/RNApage.html (many links to a variety of RNA-related sites)

Exercise 1.2: Folding the Tape: Higher Orders of Structure

As you saw in Exercise 1.1, it is not really possible to visualize a molecule's shape from a printed string of subunits. Yet this is the same type of data that a molecular biologist is confronted with after completing the biochemical sequencing of an unknown RNA molecule. Fortunately, there is a better way.

In this part of Activity 1, you will use a simple molecular model and a specific set of folding rules to discover and visualize some common structural features of RNA. You will fold simple linear models of the RNA sequences in Figure 1.3 into more complex shapes by applying folding rules.

Materials

- pipe cleaners
- colored markers
- template sheets for color-coding
- tape

Procedure

1. Each team member receives four white pipe cleaners (30.5 cm each). Take *two* pipe cleaners and make an elongated model 50.5 cm long by overlapping 11 cm of each pipe cleaner and twisting the ends together thoroughly. Repeat this procedure with your two remaining pipe cleaners to produce a second model.

2. Each team member selects a *different* sequence from Figure 1.3 and colors *two* identical pipe cleaner models of that sequence. Use colored markers and the following code to mark the 34 nucleotide bases of your sequence as colored patches on the models:

   yellow = A green = U black = G white = C

   Make each colored patch 15 mm long. To do this, lay both pipe cleaners side by side over the paper template provided by your instructor. You can now easily color both models at the same time. (It helps to tape the ends of the pipe cleaners to the template.) Be sure to color completely around each model's circumference. Double-check your colored models for coding errors before you proceed.

3. Indicate the polarity of the RNA molecule by marking the 5′-end with a black dot; the opposite end is the 3′-end. (In the real molecule, the 5′-end nucleotide would have a free phosphate group at position 5 in its ribose sugar; the 3′-end nucleotide would have a free OH group at position 3 of its ribose).

4. Distribute the models within your team as follows: two persons *each* get a copy of sequence 1 and sequence 3; two persons *each* get a copy of sequence 2 and sequence 4. Note: You can easily distinguish the different sequence models by noting that the 5′-end base is unique for each.

Figure 1.5 Folding rules. Use these rules to fold linear RNA molecules into more complex shapes.

- RNA molecules are highly *flexible*. Bases *within* the molecule can pair as follows: A pairs with U; G pairs with C.
- Molecules fold so as to *maximize* the number of paired bases.
- Paired bases create a *double-stranded stem region* in the molecule; these stem regions must contain *three or more* base pairs to be stable.
- Double-stranded stem regions can only form by *antiparallel association* of two single-stranded regions (that is, one strand must run 5′ to 3′, while the partner strand runs 3′ to 5′).
- *Single-stranded loops* can form in the molecule; they must be *three or more* nucleotide bases in length.
- Three or more bases in a single-stranded loop can pair with another single-stranded region elsewhere in the molecule. A three-dimensional (tertiary) shape called a *pseudoknot* results.

These rules are based on observations scientists have made of the folding behavior of biologically active RNA molecules.

Use the folding rules listed in Figure 1.5 to fold your two sequences. Hint: Work in parallel with the team member folding the same sequences as you, so that you can help each other.

5. **Compare your structures with those of your teammates and with those of other teams. Reject any invalid structures that do not conform to the folding rules. The valid models represent the potential three-dimensional structures dictated by your RNA sequence.**

Challenge Questions

1. **Draw a picture of each RNA's shape.**

   The shapes are illustrated in Figure T1.1 on page 37.

2. **Write a detailed description of the three-dimensional shape for your RNA sequences, indicating the structural features present. Figure 1.6 depicts some common structural elements of RNAs.**

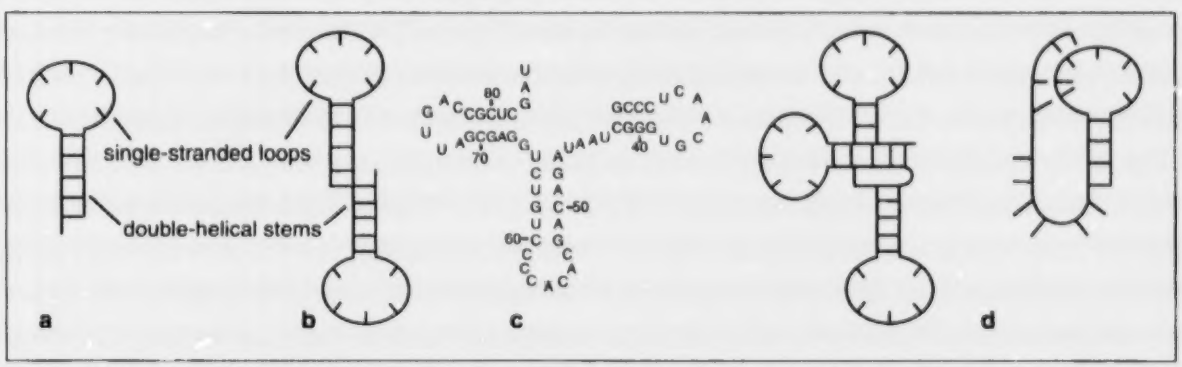


Figure 1.6 Some examples of RNA shapes and structural elements. Single molecules of RNA consisting of one (a), two (b), or three (c) stem-loop elements. (d) A pseudoknot configuration.

Figure 1.6 depicts some common structural elements of RNA. Students can use these elements to describe the structures that emerge in the RNA models.

3. **Do any of your sequences have more than one possible structure that fits the folding criteria?**

Sequence 3 has two possible structures.

FOR YOUR INFORMATION
RNA Structure: Tapes to Shapes

The shapes in Figure 1.6 represent some aspects of RNA structure, but they do not fully capture the potential for three-dimensionality in the molecule. Because RNAs perform many different tasks in the cell, a variety of different and complex shapes is required. Like proteins, complex RNA shapes arise through interactions at several structural levels (Figure 1.7).

The order of the nucleotide monomers in the tapelike *primary* (1°) *structure* of RNA or DNA forms a code, both for storing genetic information and for determining the molecule's shape. In *DNA*, two linear molecules intertwine into the familiar double-stranded helix, with complementary nucleotide bases of adjacent strands pairing by hydrogen bonds. RNA is more likely to exist as a single molecule; nevertheless, hydrogen base interactions are important here as well.

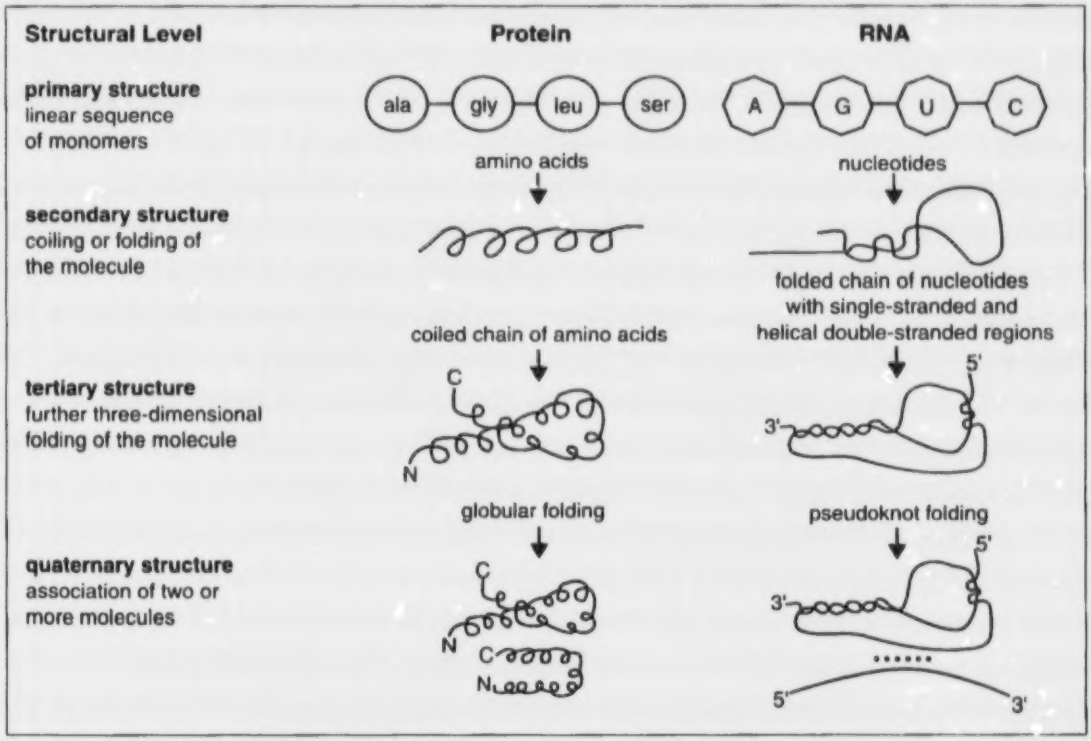


Figure 1.7 Comparison of RNA and protein structural elements.

33

FOR YOUR INFORMATION (continued)

The primary sequence of RNA is highly flexible, and thus can bend and fold on itself. This folding permits bases within the *same* molecule to approach one another and pair via hydrogen bonds. Base pairing within an RNA molecule obeys pairing rules similar to those that operate between the two strands of DNA: C pairs with G, and A pairs with U (T in the case of DNA). This folded level of structure is called *secondary* (2°) *structure*. The monomers involved in this base pairing may be widely separated in the primary sequence, but folding of the molecule brings them close enough to bond. Many shapes can result, resembling a hairpin, a stem with a large loop, even a clover leaf. Note that base pairing within an RNA molecule produces local double-stranded regions that take the shape of a helix, similar to a DNA helix. Regions not involved in base pairing remain single-stranded.

Scientists now realize that even higher levels of structure occur in RNA. Like proteins, *tertiary* (3°) and *quaternary* (4°) interactions are possible. For example, a single-stranded loop region in an RNA may fold over and base-pair with some other single-stranded region of the molecule. This forms a three-dimensional shape known as a *pseudoknot*, an example of tertiary structure in RNA. It is even possible for two different RNA molecules to interact via base pairing between their single-stranded regions, an example of quaternary structure.

Exercise 1.2: Alternative Modeling Method

In this part of Activity 1, you will use a simple molecular model and a specific set of folding rules to discover and visualize some common RNA shapes. You will fold a simple linear model of the RNA sequences in Figure 1.3 into more complex shapes by applying folding rules.

Materials

- vinyl tubing
- preprinted strips of paper
- twist-ties

Procedure

1. Each team member receives two preprinted strips of paper, each bearing a *different* base sequence listed in Figure 1.3, as follows: Two members *each* receive a copy of sequence 1 and a copy of sequence 3; the other two members *each* receive a copy of sequence 2 and a copy of sequence 4. The 3′- and 5′-ends of the molecule are indicated on the strip.

2. Each team member also receives *two* 30-cm lengths of clear vinyl tubing and 10 twist-ties. Insert each of your paper-strip sequences into a separate tube. (Allow the end tab that bears the sequence's number to protrude from the tube.) Hint: Insert the strip slowly while keeping the tube straight. This is easily done by holding the tube vertically while securing one end under your foot.

3. **Use the folding rules listed in Figure 1.5 to fold your sequences. Hint: Work in parallel with the team member using the same sequences as you, so that you can help each other. Use twist-ties to represent a few of the hydrogen bonds and to stabilize your model.**

4. **Compare your structures with those of your teammates and with those of other teams. Reject any invalid structures that do not conform to the folding rules. The valid models represent the potential three-dimensional structures dictated by your RNA sequence.**

Challenge Questions

1. **Draw a picture of each RNA's shape.**

   The shapes are illustrated in Figure T1.1 on page 37.

2. **Write a detailed description of the three-dimensional shape for your RNA sequences, indicating the structural features present. Figure 1.6 depicts some common structural elements of RNAs.**

   Figure 1.6 depicts some common structural elements of RNA. Students can use these elements to describe the structures that emerge in the RNA models.

3. **Do any of your sequences have more than one possible structure that fits the folding criteria?**

   Sequence 3 has two possible structures.

In Exercise 1.2, students create models, apply folding rules, and manipulate their models to explore the shapes of RNA base sequences. Students use simple pipe cleaner or tubing models to explore the consequences on shape of base interactions within the RNA molecule. This exercise lets students discover the following aspects of RNA shape:

- higher order (2° and 3°) shapes arise when bases interact within the RNA molecule;
- related base sequences can have related shapes;
- a given sequence may be compatible with more than one shape; and

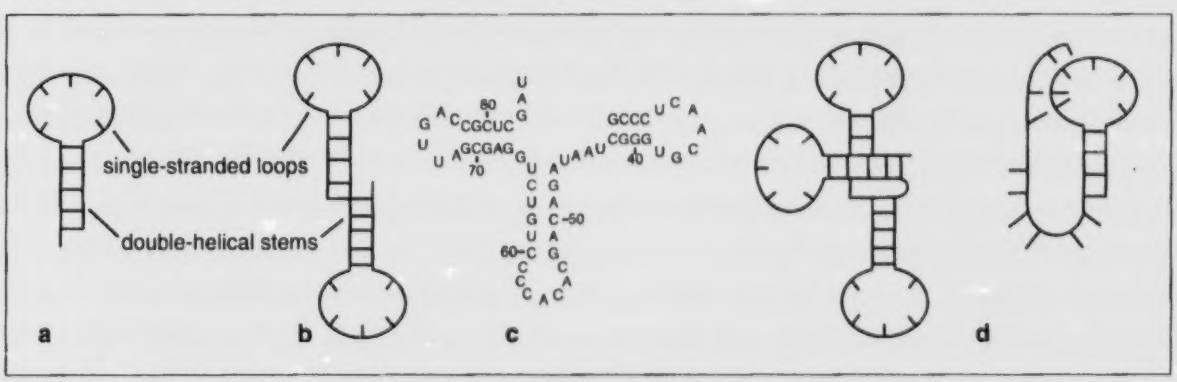


Figure 1.6 Some examples of RNA shapes and structural elements. Single molecules of RNA consisting of one (a), two (b), or three (c) stem-loop elements. (d) A pseudoknot configuration.

- certain structural motifs are common in RNAs (for example, helical stems, single-stranded loops, bulges, and pseudoknots).

Later, in Exercise 1.3, students explore more subtle concepts related to shape: molecular function depends on correct molecular shape, and some changes in base sequence (mutations) alter RNA shape and function while others do not.

Students use pipe cleaners to make the linear models of RNA sequences. The pipe cleaners are marked with four colors, coded to correspond to the four sequences of nucleotide bases shown in Figure 1.3. Students can mark the pipe cleaners or you may mark them to save class time. The color scheme follows:

yellow = A green = U black = G white = C

If students mark the models, remind them that it must be done *accurately* for the model to be useful; you might suggest that they double-check their work.

Within each group, two students work with models of sequences 1 and 3; two work with models of sequences 2 and 4. Encourage team members to collaborate as they attempt to fold their sequences together.

Note that an *alternate modeling procedure* is also provided. This procedure uses clear vinyl tubing models into which the students insert strips of paper bearing the base sequences. Both modeling procedures use the same sequences and highlight the same points, but they have different advantages and disadvantages regarding instructor preparation time and student manipulations during class. Decide ahead of time which procedure is most practical. The printed base sequences appear in the *Copymaster* section; photocopy them and cut apart the sequence strips.

After the models are ready, you may want to spend a few minutes explaining the folding rules (Figure 1.5) before the students begin folding.

Point out to students that while these modeling exercises are artificial in terms of the convenient choice of the short sequences used, real "rules" for molecular folding do indeed exist in nature; in this way, the exercises simulate real events. Once they have manipulated the models, students will discover for themselves that *some sequences are potentially compatible with more than one shape* under the folding rules. Later, during the discussion, you can point out that thermodynamic issues and the biochemistry of base pairing make certain shapes more likely to form or to be stable. Students will also see that *the same or a similar shape can result from more than one base sequence*. This situation may occur when the different sequences are genetically related or share a common function. The predicted shapes of the four RNA sequences are shown in Figure T1.1.

As an aside, you may want to do the following simple demonstration, suggested by Jacek Wower of Auburn University, to illustrate the significance of hydrogen bonding for nucleic acid stability. While a single hydrogen bond is weak, the combination of many hydrogen bonds provides the forces necessary to stabilize a stem or helix in nucleic acids. You can easily make this point by having students fold a piece of wire in half and twist it once to represent a single hydrogen-bonded base pair. Instruct students to grasp the ends of the wire and pull; it is fairly easy to unfold the wire. Now

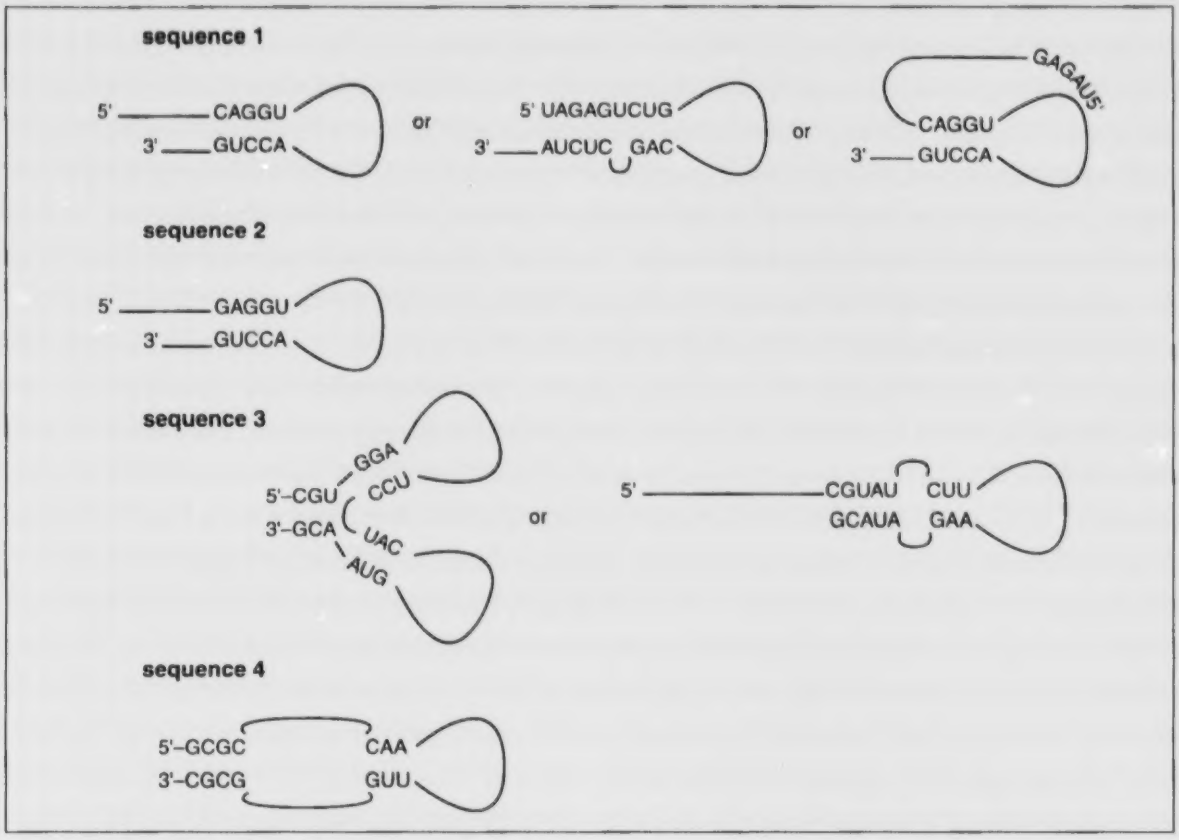


Figure T1.1 Possible shapes of RNA sequences in Exercise 1.2.

repeat the exercise by folding the wire in half and twisting several times to represent multiple base pairs. When students grasp the wire by the ends and pull, they find that the multiple associations make it difficult or impossible to straighten it.

Encourage students to see that in using models, whether simple physical models like these or sophisticated computer modeling software, they are doing science as it is done in research settings. Models are certainly useful for teaching about known molecular shapes, helping teachers more graphically communicate to students the important features of molecules and their function. But models also serve as important research tools to help scientists discover new molecular structures and functions.

After working with their models, team members should compare their RNA structures in preparation for a class discussion of possible structures of the four different sequences. Allow students to debate within their teams about the validity of various structures and to reject structures that do not meet the folding criteria. Encourage students to rework their RNA models to test some of their classmates' hypotheses regarding feasible structures. Point out that, in a sense, this debate and comparison of structures mimics the real-life process of scientists reporting their results at scientific meetings. Different research groups often determine alternative possible structures for a particular RNA sequence, and comparison of results can show errors or expand the view to include all possibilities.

Exercise 1.3a: Finding Function

Now we will relate RNA structure to function by determining which of your four RNA sequences is the biologically active structure capable of carrying out a specific simulated molecular task, in this case, binding to a particular "protein."

Materials

- **models from Exercise 1.2**
- **styrene ball**

Procedure

1. **After your team has agreed on valid models for each sequence from Exercise 1.2, arrange one set of models on a desk. Each team receives a 5-cm diameter styrene ball that simulates a small globular protein molecule. Note that the ball has several letters marked on it. Each letter represents the position of a different amino acid in the primary sequence of the protein. When you align the protein model with the RNA model that has the correct "functional" base sequence, the letters on each model will match up identically, indicating a functional fit.**

2. **Working as a team, test each of your RNA models for its ability to "bind" the "substrate protein."**

Challenge Questions

1. **For which sequences did you find a function?**

   All four of the sequences had a function, in the sense that they underwent a sequence-specific interaction with the "protein" balls.

2. **For sequences having more than one possible shape, did both shapes have the same functional ability? What does this imply about the relationship between molecular structure and function?**

   For sequences having more than one possible shape, only one of the shapes was functional in being able to bind to the ball. This implies that the relationship between molecular structure and function is quite specific.

3. **Can you think any functions performed by RNAs? List them.**

   Functions performed by RNAs include binding to other RNAs and to proteins. This kind of interaction is found, for example, in the structure of ribosomes, spliceosomes, and signal recognition particles. RNAs also function as mRNAs and tRNAs, among other functions.

In Exercise 1.3a, students apply a functional test to their models. Students attempt to align a marked, styrene "protein molecule" with their RNA models. They should see that some but not all of the RNA models "work" in binding the ball, allowing them to make clear the connection between a molecule's three-dimensional structure and its biological function. Students may already be aware that in proteins (for example, enzymes or receptors), structure is important for function.

But they may not have this concept at the forefront of their thinking, and most likely they have not had occasion to generalize the structure-function relationship to nucleic acids. The FYI essay *RNA Structure: Tapes to Shapes* expands somewhat on the general principles underlying shape determination in RNA and DNA.

You may remind students that the usefulness of a particular structure depends in part on external conditions that may change across time. You can make an analogy for structure-function selection by comparing the utility of a metric wrench with a nonmetric wrench for loosening a bolt: The key lies with the "environment," in the sense of the type of bolt to be loosened. Both wrenches could be functional, but only one will work optimally with a metric bolt. Similarly, the usefulness of a particular molecular structure depends on the nature of the job that needs to be done, such as the particular substrate to be bound.

Directions for Marking Styrene Balls

FOR THE ALTERNATIVE, VINYL TUBING MODELS

Use styrene balls with a 5-cm diameter.

Figure T1.2 shows the approximate markings for the four balls used in the alternative exercise (preprinted RNA sequences with vinyl tubing). The letters on the balls stand for the particular bases in each RNA sequence that align with the ball at those points. Note: To avoid confusion, be sure to explain to students that the letters on the ball do not represent nucleotide bases in the protein, but simply mark the locations in the protein of amino acids that align with these particular base locations in the RNA model.

Two views of a ball are shown in the figure, which represent views of the "front" and "back" of the ball *before* and *after* rotating the ball 180° horizontally around its equator. The letters are centered approximately 12 mm apart.

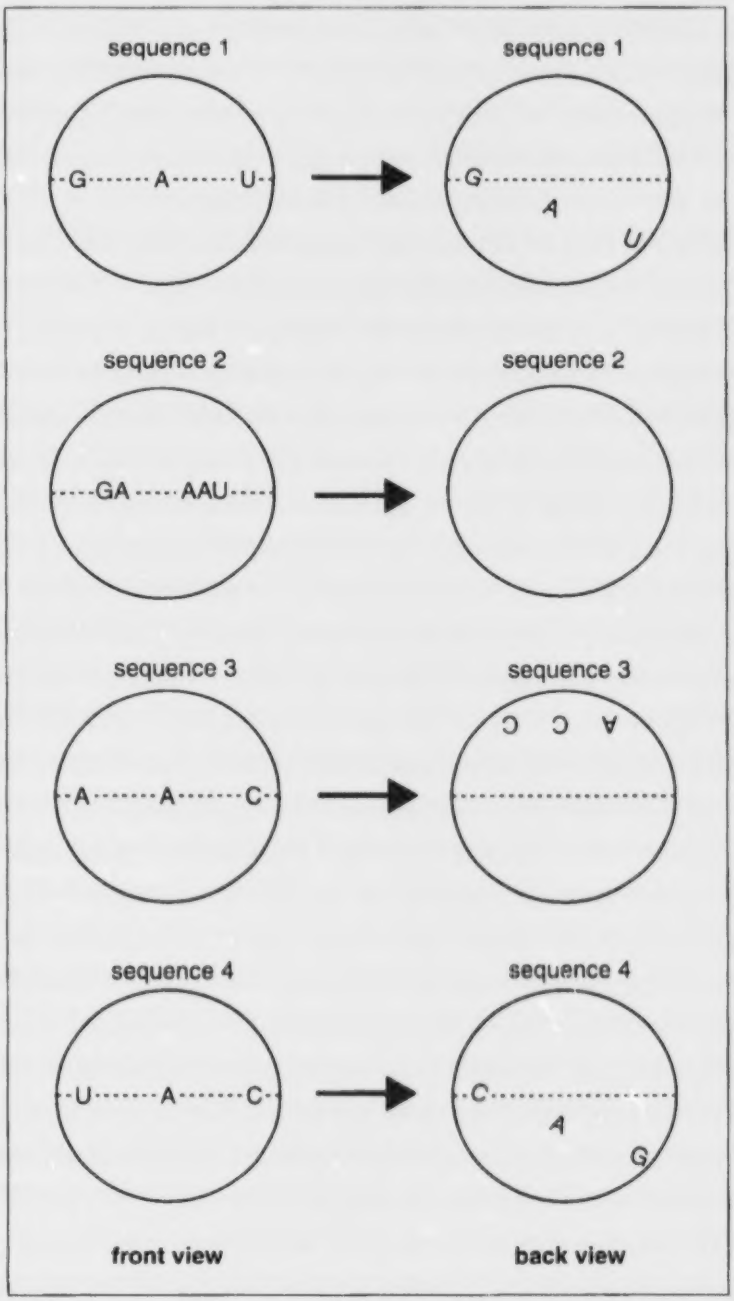


Figure T1.2 Location of marks on the styrene balls.

FOR THE PIPE CLEANER MODELS
In place of the letters shown in Figure T1.2, draw adjacent line segments (each 12 mm long) of the appropriate color (use pencil on the white ball to represent a white line segment).

It is a good idea to assemble one of each RNA model and check the positioning of the markings before making the rest.

Exercise 1.3b: Effects of Mutation

Now you will determine whether a mutation (change) in the primary base sequence of the RNA molecule will alter the shape and function of the molecule.

Materials

- **models from Exercise 1.2**
- **styrene ball from Exercise 1.3a**
- **dull tape**
- **colored marking pens**

Procedure

1. **Your instructor will give you information showing a mutation in your primary RNA sequences. Using your model, determine whether the mutation has any effect on the structure of your RNA molecule. Hint: You can cover the section of the model to be mutated with dull tape and use a colored marking pen to change the base to fit the mutation data.**

2. **Test each of the mutated sequence models for function using the "protein" ball from Exercise 1.3a.**

Challenge Questions

1. **Write a brief description of any structural effects that the mutations had on your model RNAs, such as "It changes the length of a loop" or "It eliminates a double-stranded stem." Indicate in your description which RNA sequence (1–4) you are describing and the base changes involved.**

2. **Do the mutations alter the function of the sequence(s) in the "ball test"? How?**

For Questions 1 and 2, students should answer that these mutations will produce the following effects in the models:

- Sequence 1 will no longer be able to form a pseudoknot tertiary structure and will lose its function in the ball test.
- Sequence 2 will acquire the ability to form a pseudoknot structure and will gain function in the ball test.
- Sequence 3 will retain its shape but will lose function in the ball test.
- Sequence 4 will retain its shape but will lose function in the ball test.

3. **After reading the FYI essay *Molecular Selection in Real Time*, in Activity 3, discuss with your teammates any general similarities between the approach to RNA structure and function that you employed in Exercises 1.2 and 1.3a and the recently developed laboratory approach called *in vitro* nucleic acid selection.**

In Exercise 1.3b, students explore the effects of mutation on structure and function. Have the students "mutate" their models by covering specific bases with dull tape and writing new bases on the tape to represent a base change. Leave the tape blank to represent a base deletion. Instruct the students to make the following specific changes in sequences 1–4:

- sequence 1 - delete the first three bases from the 5′-end
- sequence 2 - change bases 2–5 at the 5′-end from UAUA to **AGAG**
- sequence 3 - change bases 9 and 10 at the 5′-end from AC to **GU**
- sequence 4 - change bases 7–9 from UAC to **AGU**

Exercise 1.4: Using the Computer to Determine RNA Structures

Having now spent some time deciphering the structures of quite short RNA molecules, you can appreciate how complex base pairing is in the more typical molecules isolated from organisms, which are hundreds to thousands of nucleotides in length. Here is where the computer is put to good use, rapidly determining the alternative shapes that a molecule is capable of assuming and calculating which of these shapes is energetically the most likely.

We will use an RNA folding program called *mfold*, which is available on the World Wide Web.

Materials

- **access to the World Wide Web**

Procedure

Access the mfold program at http://www.ibc.wustl.edu/~zuker/rna/form1.cgi. Apply the program in turn to each of the sequences in Figure 1.3 that you folded. The program is easy to use, and you can simply follow the on-screen instructions to analyze your sequences. Following are some points to keep in mind as you use the program:

1. **The program analyzes only one molecule at a time; you must go through the entry and analysis process for each sequence. Once the program has analyzed and displayed a molecule, you can simply delete that sequence from the data entry field and enter a new sequence to be analyzed.**

2. **Enter sequences beginning with the 5′-most base.**

3. **Use the default parameters displayed on screen (for example, *the molecule is linear* or *fold at 37 degrees*), with the following exceptions: choose *image resolution low* and *structural format: bases*.**

4. **Once you have entered the sequence and analysis parameters, scroll to the bottom of the page and click on *Fold RNA*. Folding and display typically require 1–2 minutes, depending on how busy the server is.**

5. **When the *Output* page appears, scroll down to the links labeled *sequence*. When more than one shape is possible, a numbered *sequence* link will appear for each. Click on any one of the numbered sequence links to open the page displaying the folded RNA structures.**

Challenge Questions

1. **Did your predictions about the shapes of the sequences match those of the computer program?**

   The computer-generated shapes should match those modeled by the students.

2. **Did the program reveal more than one shape for any of the molecules? If so, which shape is more likely on energetic grounds?**

   Answers will vary.

In Exercise 1.4, students use an RNA folding program to verify the structures of their model sequences. Several such programs are available, varying in sophistication and ease of use. We have chosen an easy-to-use program called *mfold*, which is available on the Web from Mike Zuker's lab at Washington University in St. Louis. Students can access the analysis page of mfold at http://www.ibc.wustl.edu/~zuker/rna/form1.cgi. (Please note that this address is active at the time of printing this module; addresses change from time to time, so you may want to use a search engine to find additional sites.)

Analysis Questions

1. **Review your answers to each Challenge Question and discuss the connection between structure and function for RNA.**

2. **If an RNA molecule functions as a catalyst, how might the folding pattern be important for its activity?**

3. **In a population of related molecules, could an altered (mutant) structure ever have an advantage? Explain.**

4. **Do you think molecules (or molecular function) can be subject to natural selection? Explain.**

5a. **Compare the consequences for a species between two changes potentially affecting its RNA structure across time: (1) a base alteration that might occur randomly in the RNA after it has been synthesized *versus* (2) a random mutational change that can be traced to the template for the RNA (such as the DNA gene that encodes it).**

5b. **Would your answer in Question 5a change if the RNA molecule itself were capable of self-replication? Explain your response.**

Extension Exercises

POLYMERS CAN ACQUIRE FUNCTION

Both short oligomers containing only a few monomer units (*oligo* is Greek for few) and longer polymers (*poly* is Greek for many) can fold into shapes that have potentially useful functions (Activity 1). But longer molecules can generally fold into more interesting shapes with a greater potential diversity of functions. In this context, it is worth noting how an irregular biological polymer such as RNA differs from a regular crystal such as sodium chloride (table salt). In a crystalline table salt, alternating sodium and chloride ions are lined up in a monotonously repeating pattern, each layer of the three-dimensional crystal just like the others. In contrast, the linear sequence of A, G, C, and U (adenine, guanine, cytosine, and uracil) in RNA is virtually limitless, allowing the molecule to fold up into a virtually limitless variety of shapes. The following exercises illustrate why long, complex RNA molecules are more likely to generate interesting shapes and catalytic functions than short RNA molecules.

EXERCISE 1.5 LONG POLYMERS ARE MORE VERSATILE THAN SHORT POLYMERS

Divide students into four groups. Each group writes down four random RNA sequences that are 24 monomer units long ("24-mers"). The groups then cut their sequences into subsequences of different lengths: The first group leaves the 24-mers intact; the second group cuts the 24-mers into 12-mers, the third group into 8-mers, and the fourth group into 4-mers. Each group assembles the sequence or sequence fragments into higher order structures based on the same rules used in Exercises 1.1 and 1.2 (G pairs with C, A pairs with U, and four consecutive base pairs are required to build a stable stem). The goal is to build an RNA molecule with a useful "function": the ability to be hung on a peg in the wall.

It will quickly become apparent that a random 24-mer can often be bent into a partially self-complementary hairpin, but 4-mers are unable to form anything other than short, paired segments that are useless for the desired purpose. The 8-mers are intermediate; a loop can occasionally be formed, but it requires luck and a lot of imagination. The lesson is that longer sequences can form functional molecules more readily than shorter ones. The first self-replicating RNA molecules would therefore have been subject to strong selective pressure to become longer. This in turn would have selected for more accurate replication, because the chances of making a fatal replication error would increase with the length of the template molecule.

EXERCISE 1.6 MONOMERIC BUILDING BLOCKS ARE MORE VERSATILE THAN OLIGOMERIC BUILDING BLOCKS (2-MERS, 3-MERS, 4-MERS, AND SO ON)

The goal of building an RNA structure that can be hung on a wall peg remains the same. However, this time the groups can build as long a polymer as needed to generate a looplike structure. Each group chooses a total of 24 monomer units from containers that you prepare; however, the first group chooses from a container of monomers, the second group from a container of dimers, the third group from a container of trimers, and the fourth group from a container of hexamers. The monomeric or oligomeric building blocks can be linked in any order (use dull tape) but cannot be cut apart into the component monomer units. The group having access to

monomers will quickly build a loop with the desired "function"; the group with dimers will probably form a loop; the group with trimers may succeed after considerable work, and the group with hexamers is almost certainly be doomed to failure. Groups may draw additional oligomers (for a total of more than 24 monomers) to see whether a greater variety of oligomers is helpful. The lesson is that it is easier to build novel structures when the sequence is entirely unconstrained; when dimer, trimer, or hexamer sequences are fixed, the number of possible rearrangements is limited. A good analogy would be to ask how well one could play Scrabble if the individual letter tiles were taped together in pairs or triplets.

Activity 2
RNA Catalysis

Activity at a Glance

Exercise 2.1a, b, c: Students use pipe cleaner models to simulate the autocatalytic self-splicing reaction of a group I intron and the nuclease activity of a hammerhead ribozyme.

Exercise 2.2: Students carry out *in vitro* transcription of RNA and study the kinetics of group I intron self-splicing in the laboratory.

Concepts

- RNA has catalytic activity as well as the ability to encode information.
- The catalytic activity of ribozymes, as well as enzymes, requires the appropriate structural chemistry and three-dimensional molecular shape.
- Catalytic RNAs can act on themselves or on other molecules.
- The kinetics of ribozyme-catalyzed reactions can be studied.

Focus

Students are challenged to apply the concepts of molecular structure from Activity 1 to the function of catalysis by RNA. In addition, students actively investigate RNA function in the laboratory.

Connections to Lecture Topics

Use Activity 2 to introduce RNA or enzymes, or with lectures on molecular structure-function relationships.

Estimated Time

Approximately 2 hours in class and 2 laboratory periods over 2 days

Materials Preparation

For each team of four students, provide

- 3 white pipe cleaners (also called *chenille stems*), 27 cm long. Directions for making pipe cleaner models are given on the template for Exercise 2.1.
- paper model of guanosine molecule (see *Copymasters*)
- paper model of an OH group (see *Copymasters*)
- scissors
- tape

See annotation and reagent listings in Exercise 2.2 for laboratory materials.

BEST COPY AVAILABLE

ANNOTATED STUDENT ACTIVITY

Introduction

When someone asks you, "What does RNA do?" you probably think of messenger RNA (mRNA), which is copied from a gene and communicates the information needed to make a protein. In this familiar role, RNA is doing a job similar to DNA, storing genetic information. This information-storage function is analogous to the way a compact disc (CD) stores musical information. To use or copy the information in a CD, however, you need a form of technology capable of performing these functions: a CD player. However, in terms of this analogy, RNA is a surprising molecule. It's a bit like a music CD that needs no CD player and can play itself. In other words, RNA has dual functions: Not only can it *store* genetic information, but in some cases it can also *transfer* and *use* this information by acting as a biochemical *catalyst*.

A catalyst is a molecule that increases the rate of a chemical reaction. Catalysts are essential to living systems because they allow biochemical reactions to occur fast enough to be compatible with life. Biochemical catalysts speed up reactions by physically binding the reacting molecules, bringing them near one another, and orienting them in just the right way to make the reaction more likely to proceed. Chemists describe this "matchmaker" function of the catalyst as a lowering of the *activation energy* of the reaction, like lowering an energy "hill" that reactants must get over before they can continue along the reaction path. The same reaction could happen without the catalyst, but it might take years instead of milliseconds to complete. Biochemical catalysts also make reactions more specific by bringing together particular molecular reactants.

The most familiar biological catalysts are enzymes. Enzyme catalysts are made of protein, and today they catalyze the vast majority of biochemical reactions, from synthetic reactions like making nucleic acids and other proteins to degradative reactions like releasing energy from sugars and fats.

Some RNA molecules also are catalysts; they are called *ribozymes*. As catalysts, ribozymes appear to function by physical mechanisms similar to those of protein-based enzymes. However, the types of reactions catalyzed by ribozymes are limited. In nature, known ribozymes primarily catalyze reactions on other nucleic acids and, very likely, some key reactions involved in protein synthesis. No doubt other ribozyme-catalyzed reactions are yet to be discovered. In the laboratory, scientists recently synthesized several novel RNA and single-stranded DNA molecules not known to occur in nature that also act as catalysts.

It may surprise you to learn that RNA (and even single-stranded DNA) can be a catalyst; this fact certainly surprised scientists when it was first discovered. At first glance, RNA appears to have little in common with proteins to account for its catalytic ability. On closer inspection, however, some critical similarities appear: Both proteins and RNA have subunits containing chemically reactive groups of atoms (for example, the imidazole group found in certain amino acids and nucleotides), and both have complex shapes. A key

point: Closely related objects (like RNA and DNA), and even not-so-closely related objects (like RNA and protein), that share key features of structure and shape can frequently perform similar functions.

In living cells, RNA molecules carry out a variety of important tasks, some catalytic and some not. For example, RNA participates in the replication of DNA, functions as the genome of many viruses, and participates in the cellular localization of newly made proteins. Another important role is in the maturation of mRNAs. mRNAs are copies of the genetic information encoded in DNA. But before an mRNA copy can be used to direct protein synthesis, it must undergo a biochemical maturation process that removes noncoding base sequences, known as *introns*. In eukaryotic cells, intron removal is accomplished by several small RNAs plus some proteins that together make up an mRNA "splicing machine"called a *spliceosome*. A few RNA introns, however, are so versatile that they can actually splice themselves out of the precursor RNA with no help from proteins. The surprising discovery of these self-splicing RNA introns by researcher Thomas Cech in 1982 first revealed RNA's catalytic ability and earned Cech a Nobel Prize.

Once a mature mRNA is formed, the subsequent steps in the synthesis of proteins rely heavily on several additional RNA molecules, from the *transfer RNAs* (tRNAs) that carry the amino acids, to the *ribosomal RNAs* (rRNAs) that, together with a variety of protein molecules, make up the ribosome itself. You recall that the ribosome is the cellular structure where amino acid monomers are linked together using the directions encoded in the mRNA to form proteins. About 80 percent of the RNA in a cell is ribosomal, and there is recent evidence to suggest that one of these ribosomal RNAs (the 23S rRNA) is in fact a ribozyme catalyst that chemically couples the amino acids during protein synthesis.

Several classes of naturally occurring ribozyme RNA have been identified in nature, and hundreds of individual examples have been found in a wide variety of living cells (Figure 2.1). Still other ribozymes have been constructed in the laboratory by conducting *molecular evolution in a test tube*. In this process, scientists start with a large, random population containing many different laboratory-synthesized RNA sequences. The researchers select from this population only those RNAs that have particular, desired functions, such as the ability to bind a particular target molecule or catalyze a particular reaction. These synthetic molecules reveal many previously unsuspected capabilities of nucleic acids, including the ability of RNA to make partial copies of itself. We will explore the important topics of *in vitro* molecular evolution and RNA self-replication in Activity 3.

You can find out more about RNA catalysis and other newly recognized cellular roles of RNA in the references listed in Additional Information. Particularly useful is a somewhat advanced book, *The RNA World*. A set of videotaped lectures about RNA by Nobel Prize winner Thomas Cech is also very informative.

Figure 2.1 Naturally occurring ribozymes.

Class	Size	Reaction	Source
group I intron	large: 413 NT in *Tetrahymena thermophilia*	intron excision	eukaryotes, eubacteria, and viruses
group II intron	large: 887 NT in yeast mitochondria	intron excision	eukaryotic organelles and eubacteria
RNase P	large: 350–410 NT	hydrolytic endoribonuclease	RNA subunit of eubacterial RNase P
hammerhead	small: 31–42 NT (enzyme strand can be 16 NT)	RNA cleavage	viral satellite RNA in plants, viroids, and newt satellite DNA
hairpin	small: 50 NT (minimum sequence)	RNA cleavage	(-) strand satellite RNA of tobacco ringspot virus
hepatitis Delta virus (HDV)	84 NT (required)	RNA cleavage	HDV
Neurospora VS RNA	881 NT (164 sufficient)	RNA cleavage	*Neurospora* mitochondria

Additional Information

Cech, T.R. (1998). Structure and mechanism of the large catalytic RNAs: Group I and group II introns and ribonuclease P. In Gesteland, R.F., & Atkins, J.F. (Eds.), *The RNA world* (pp. 239–269). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Landweber, L.F., Simon, P.J., & Wagner, T.A. (1998). Ribozyme engineering and early evolution. *BioScience*, 48(2): 94–103.

Howard Hughes Medical Institute Holiday Lecture on Science. (1995). *The Double Life of RNA* [videotape set]. A four-tape series of lectures by Nobelist Thomas Cech, discussing his discovery of ribozymes and other important functions of RNA. Summaries of the lectures can be found online at http://www.hhmi.org/.

Challenge Questions

Read the Introduction, then discuss any general similarities between catalysis by proteins and nucleic acids.

Students should identify the following similarities:

- both act on substrates;
- both rely on precise, three-dimensional shape to accomplish catalysis; and
- both remain unaltered during the reaction (self-splicing introns are an exception because they act on and modify themselves).

The Introduction acquaints students with the notion that RNAs have a dual function as information storage molecules and as functional agents within the cell. Students are introduced to the concept that some RNAs have catalytic functions. You

can build on these recently recognized functions of RNA by explaining the more familiar functions of catalysis for enzymes and of information storage for DNA. The idea that some naturally occurring RNA molecules are catalysts may be new to many of your students. Also introduced is the technique of *in-vitro* selection (that is, molecular evolution in a test tube), which helps students understand that molecules, too, can be subject to natural selection. The important topic of *in-vitro* selection is dealt with directly in Activity 3, *RNA and Evolution*.

Exercise 2.1a: Modeling the Hammerhead Reaction

In this activity, you will use pipe cleaner models to model the catalytic activity of two RNA molecules: a simple catalytic RNA called the hammerhead ribozyme and the first catalytic RNA to be discovered, the self-splicing group I intron.

First you will examine the catalytic activity of the so-called hammerhead ribozyme. The name derives from the superficial resemblance of this RNA's secondary structure, as usually depicted, to a carpenter's hammer. Hammerhead RNA sequences are embedded in the RNAs of many viruslike agents that infect plants. The RNA genome of these so-called viroids and virusoids is replicated in the form of a long precursor transcript made up of several tandemly repeated genomes plus repeats of the smaller hammerhead sequence. The hammerhead sequences catalytically cut themselves out of the long precursor transcript, in the process cleaving it into many individual genome-length segments. The hammerhead RNA sequence is thus an *endoribonuclease*, a catalyst able to cleave RNAs internally.

Materials

- **pipe cleaner model of hammerhead ribozyme**
- **pipe cleaner model of substrate RNA**
- **scissors**
- **tape**

Procedure

The pipe cleaner model with four clusters of labeled bases represents a hammerhead RNA sequence that has been genetically engineered to interact with an external RNA substrate. (An external substrate is an RNA molecule other than the one in which the hammerhead resides.) The other labeled pipe cleaner represents the substrate sequence for this hammerhead ribozyme.

1. **Use the base sequences in Figure 2.2 to locate and mark the 5′-end of each pipe cleaner model (use a black ink dot or marking of your choice).**

2. **Use the sequences on your hammerhead and substrate pipe cleaner models to fold them into the appropriate secondary structures.**

3. **Identify a base-pairing interaction *between* the hammerhead and substrate molecules.**

BEST COPY AVAILABLE

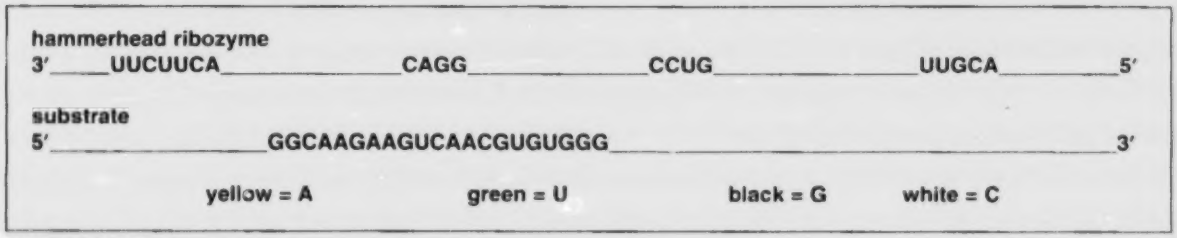


Figure 2.2 Base sequences.

4. **Locate the enzymatic cut site in the substrate. This site has the general sequence NUH, where N can be any nucleotide, and H can be A, C, or U.**

5. **Cut the substrate on the 3′ side of the base you identified as H.**

Challenge Questions

1. **Describe what this reaction accomplished. Draw the resulting molecules.**

   The hammerhead reaction resulted in an endonucleolytic cleavage of the substrate RNA.

2. **Suggest a practical application for such an endonuclease.**

   This type of enzymatic activity could function in the cell as a general-purpose RNA nuclease.

As you can see, formation of the hammerhead ribozyme's catalytic site depends on the molecule first folding into the correct structure.

Recall that the naturally occurring hammerhead sequences in the plant viroids and virusoids typically cut only the RNA molecule in which they are embedded. In contrast, hammerhead sequences like the one you just modeled, which can cleave *external* substrates, have for the most part been fashioned in the laboratory through genetic engineering. These engineered ribozymes function like protein enzymes because they react with many copies of an external substrate in a multiple turnover process. To develop a potential form of gene therapy, researchers are attempting to design RNA-cutting RNAs like these, which can recognize and cut either foreign RNAs (such as the HIV viral genome) or mutated cellular RNAs (such as the altered RNAs that can promote cancer). The hope is that when these ribozymes are incorporated into affected cells, they will specifically eliminate the offending target RNAs.

Exercise 2.1b: Modeling the Group I Intron Reaction

Next, you will model the first catalytic RNA discovered, the self-splicing group I intron. This RNA was first found in 1982 by Thomas Cech in the nuclear ribosomal RNA of the protozoon *Tetrahymena thermophilia*. Introns of this type also occur in other eukaryotes as well as in prokaryotes and viruses. The group I intron RNA is said to be *autocatalytic* because it acts on itself. The intron resides within a larger precursor rRNA molecule and catalyzes a

two-step reaction that dramatically alters its relationship to the larger molecule. Like other functional RNAs, group I introns must first fold into a precise secondary structure by internal base pairing. Proper folding enables the intron to recognize and react with particular bases at the *splice sites* within the larger RNA.

Materials

- pipe cleaner model of RNA
- paper model of guanosine molecule
- paper model of an OH group
- tape

Procedure

The pipe cleaner model provided represents a precursor rRNA containing a group I intron sequence flanked on either side by exon sequences. The 5′- and 3′-ends of the molecule are indicated, and several of its nucleotides are color-coded as follows:

yellow = A green = U black = G white = C

The following key bases are marked in the model:

5′___CCCUCUA_____UUUA_____AGAGGG_____GU_____3′

1. Locate the following landmarks in the model:
   - The 5′-most exon, which *ends* with the base sequence CCCUCU.
   - The 3′-most exon, which *begins* with and includes the U of a marked G-U pair.
   - The self-splicing intron is everything in between these exons.
2. Use the sequence on the model to form an appropriate secondary structure.
3. Simulate a nucleophilic attack and cleavage by the 3′-OH group of the free guanosine. This cleavage occurs at the 5′ exon-intron boundary of the RNA (at a PO_4^- group not shown in the model). Simulate cleavage by cutting the pipe cleaner *between* the U and the A at the 5′ exon-intron boundary.
4. After cutting, attach a paper OH group to the 3′-most base of the 5′ exon; this creates a reactive OH group at the end of the 5′ exon. Next, attach the paper guanosine by its 3′-OH to the 5′-most end of the intron.
5. Simulate the second step of the self-splicing reaction by using the 5′ exon's exposed OH to attack and cleave at the 3′ exon-intron boundary. Simulate this by making a cut between the G-U pair at this boundary. After cutting, transfer the 5′ exon's exposed OH group to the 3′-most end of the intron.
6. Bring the 3′-end of the left exon next to the 5′-end of the right exon and tape them together.

Challenge Questions

1. **Describe what these reactions accomplished. Draw the resulting molecules.**

   The intron excises itself from the larger RNA and, in the process, ligates the 5′ and 3′ exons together.

2. **What is the base sequence at the junction of the spliced exons?**

   5′...CCCUCUU...3′.

3. **Where did the guanosine end up?**

   The guanosine is incorporated at the 5′-end of the intron itself.

4. **Which catalytic ability does the group I intron possess that is lacking in the hammerhead? in the ribozyme?**

   Ligation ability. Both molecules can cleave, but the intron can ligate as well.

Exercise 2.1c: The Group I Intron RNA Can Go a Step Further

Procedure

After removing itself intact from the larger rRNA precursor, as you have just demonstrated, the group I intron often continues its cleavage-ligation activity by reacting on itself internally. In this secondary, follow-up reaction, a reactive nucleophile group at the 3′-end of the intron (the chemical nature of this group should be familiar by now) cleaves the intron at a particular purine-pyrimidine junction.

Use the reactive 3′-OH of the intron as a nucleophile to carry out another cleavage-ligation reaction. Remember, the nucleotides are color-coded as follows:

yellow = A green = U black = G white = C

1. **Locate the only remaining purine-pyrimidine junction marked in the intron model.**

2. **Simulate the intron's cleavage of itself by cutting it between the purine and pyrimidine bases.**

3. **Transfer the intron's 3′-OH to the pyrimidine base newly exposed at the end of the shorter fragment.**

4. **Simulate ligation within the intron by taping its free ends together.**

Challenge Questions

1. **What is the end result of this reaction? Draw the resulting molecules.**

   The intron removes from within itself a short linear fragment of sequence: 5'GA.....UUU-OH3'.

2. **Where does the original guanosine molecule end up?**

   The original guanosine ends up as the 5'-most G of the shorter fragment.

As you have seen, the group I intron is self-splicing, catalytically removing itself from within the larger precursor rRNA in which it resides. This self-splicing occurs by a two-step reaction sequence: a *cleavage* reaction, which excises the intron from between the two coding exons that flank it, followed by a *ligation* reaction, which splices the exons together to create the mature, and now smaller, rRNA. These cleavage-ligation reactions require no source of chemical energy, such as ATP or GTP, because they simply entail the replacement of one phosphodiester bond with an energetically equivalent one (such reactions are termed *transesterifications*). All that is required *in vitro* for self-splicing by the group I intron is free guanosine (or any of its phosphorylated derivatives: GMP, GDP, or GTP) plus a divalent cation (typically Mg^{++}). Group II introns, found in cyanobacteria and eukaryotic organelles such as mitochondria and chloroplasts, have simpler requirements for their self-splicing and do not need a free nucleotide.

Naturally occurring self-splicing introns react only with themselves and thus not with external substrates. This self-limited reactivity stands in contrast to the repeating action of protein enzymes, which typically react with many copies of an external substrate in a repetitive turnover process. Some

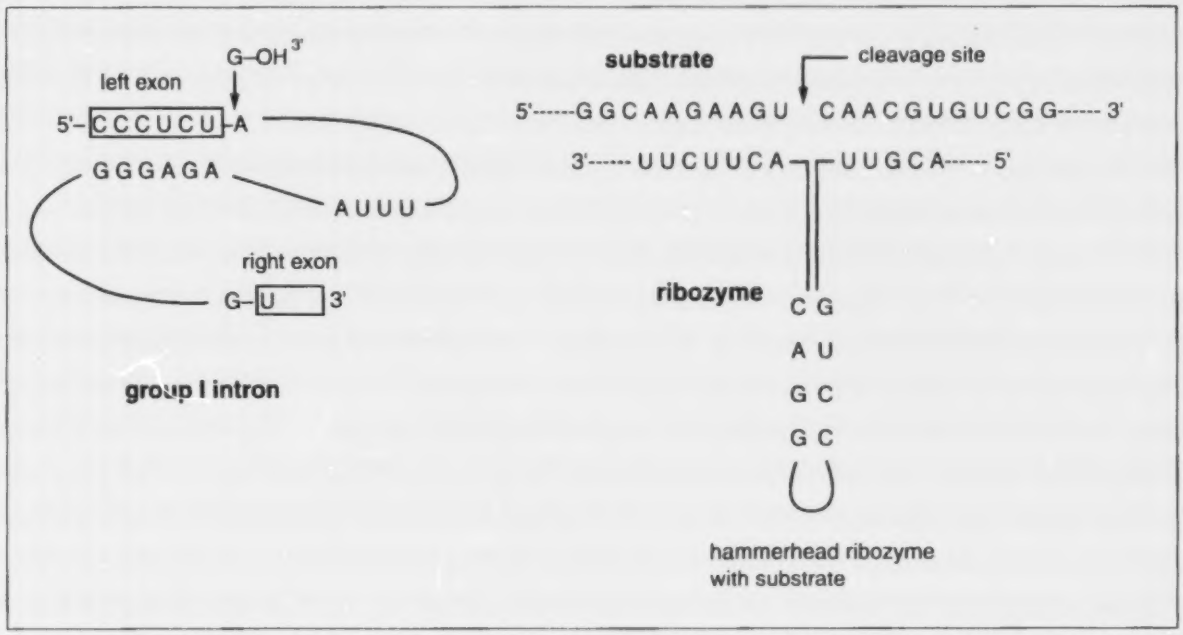

G–OH 3'
left exon
5'– C C C U C U – A
G G G A G A
A U U U
right exon
G– U 3'
group I intron
substrate
cleavage site
5'------G G C A A G A A G U C A A C G U G U C G G------ 3'
3'------ U U C U U C A U U G C A------ 5'
ribozyme
C G
A U
G C
G C
hammerhead ribozyme
with substrate

Figure T2.1 Folded molecules. Folded pipe cleaner models of the group I ribozyme and the hammerhead ribozyme with substrate.

ribozymes of the nonintron type (such as ribonuclease P, which participates in the cleavage and maturation of transfer RNAs), are in fact capable of reacting with multiple copies of an external substrate.

The folded molecules are shown in Figure T2.1.

Exercise 2.2: A Laboratory Study of the Self-Splicing RNA from *Anabaena*

In this laboratory experiment, you will carry out in the test tube two important reactions involving RNA. The first is transcription, a classic reaction in which a DNA sequence is copied into a complementary RNA sequence. The RNA sequence that you will transcribe contains a catalytic group I intron like the one you modeled in the previous exercise. In the second part of this experiment, you will examine the autocatalytic self-splicing activity of this RNA. The discovery of catalytic RNAs, or ribozymes, of which the group I intron is just one example, resulted in a joint Nobel Prize for Thomas Cech and Sidney Altman in 1989.

In most eukaryotic genes and some prokaryotic and phage genes, the genetic information is not colinear with the chromosomal DNA sequence. Rather, the information is interrupted by noncoding sequences, some many hundreds of bases long. Chromosomal DNA is therefore a mix of coding regions, called *exons*, and noncoding regions, called *introns*. During transcription, the exon and intron sequences are transcribed together into a primary RNA transcript. In order to be translated into protein form, this primary transcript needs to be processed into a mature RNA. Processing involves catalytic removal of the introns and the simultaneous splicing together of the exons. Such *post-transcriptional processing* is not confined to mRNA production, but is also found in the production of functional rRNA and tRNA.

Four biochemical classes of spliceable intron have been identified. Some of these need accessory proteins and/or an energy source to catalyze splicing efficiently, but the so-called group I and group II introns require neither: They are self-splicing. All that these RNAs require for self-splicing is a guanine nucleotide cofactor (guanosine, GMP, GDP, or GTP) and a divalent cation (Mg^{2+}). The RNA you will use is a pre-tRNA from the cyanobacterium *Anabaena* and contains a group I intron.

You will transcribe this autocatalytic RNA *in vitro* from an engineered plasmid (pAtRNA-1) that contains a cloned copy of the $tRNA^{Ile}$ DNA gene sequence (PCC7120) from *Anabaena*. This tRNA-encoding DNA was cloned into a commercial plasmid (pBS-) downstream from a copy of the T7 promotor. The promotor provides a transcription start site for the enzyme T7 RNA polymerase, which you will use to copy the insert DNA into a functional RNA *in vitro*. The Cech lab further engineered the insert DNA at its 3′-end to include an *EarI* site, a particular restriction endonuclease site. Cutting open the circular plasmid with the *EarI* enzyme at this site linearizes the plasmid and provides a transcription termination site for the T7 RNA polymerase. This type of termination is called *run-off transcription*, as the polymerase literally falls off the end of the linear DNA when it reaches the *EarI* site. The

BEST COPY AVAILABLE

linearized plasmid will yield a primary RNA transcript of 334 nucleotides in our experiment.

The self-splicing reactions that form the mature $tRNA^{Ile}$ involve both cleavage and ligation. A schematic of the reaction is shown in Figure 2.3. Note that it is a two-step mechanism in which different-sized fragments are produced during each step. Initially, we have the primary transcript with a length of 334 bases. In the first reaction step, exogenous guanosine functions as a nucleophile to attack and cleave the intron at its 5′-end. The resulting cleavage of the RNA at this 5′ exon-intron boundary yields a 296 NT intermediate (plus a small 40 NT fragment representing the 5′-end of the exon). The second step, attack of the 5′-end of the exon on the 3′ intron-exon boundary, cleaves the RNA at this boundary, ligating the 5′- and 3′-ends of the exon into a mature 85 NT tRNA and releasing the free 250 NT intron. The overall result of the cleavage-ligation reactions is the release of the 250 NT intron plus the mature 85 NT tRNA molecule.

These reactions are simple transesterifications (involving the replacement of one phosphodiester bond with an energetically equivalent one) and thus

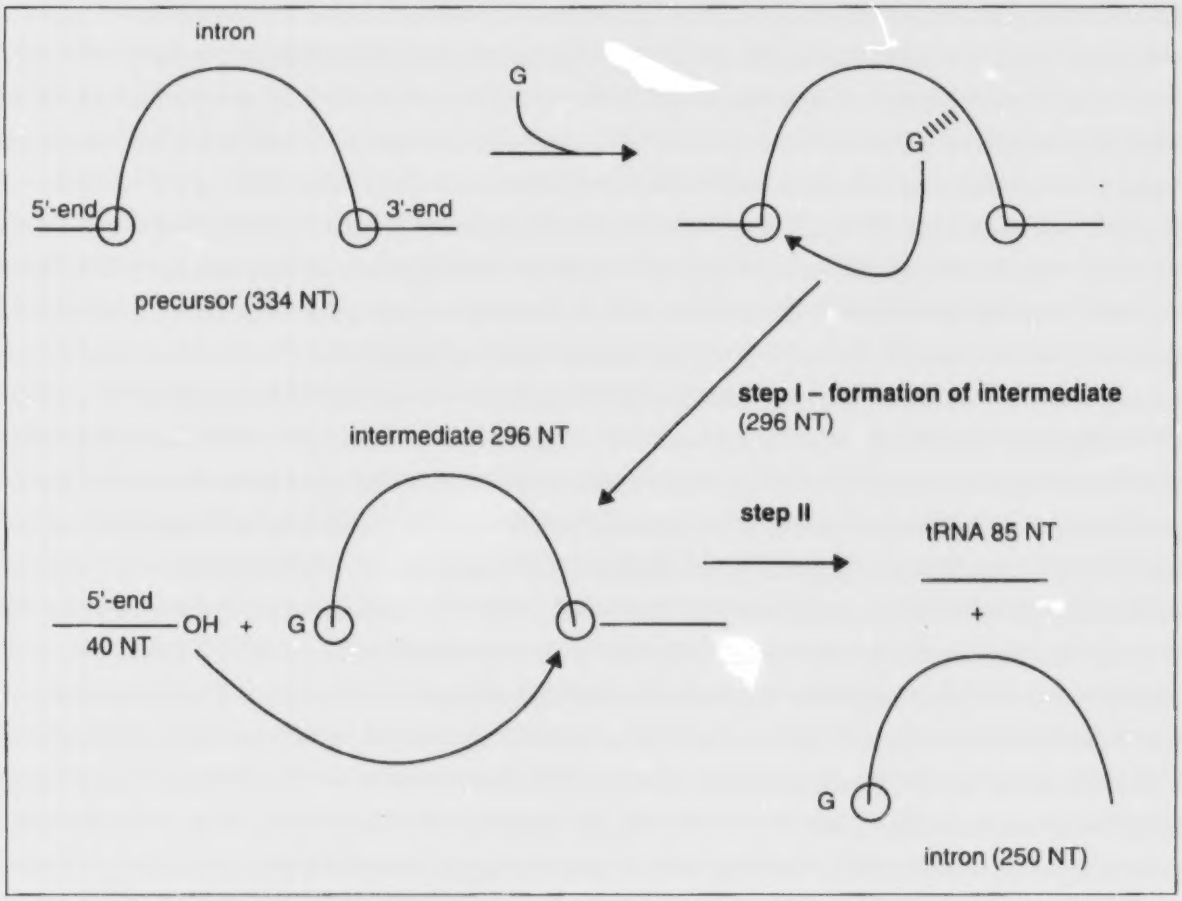


Figure 2.3 Self-splicing reaction of the *Anabaena* group I intron.

require no source of chemical energy. We can study the kinetics (time-dependent changes) of such a two-step reaction and determine which of the steps is slower (rate limiting). This can be done by simply observing the pattern of splicing products. If the first step is faster, then the intermediates made in the first step (296 NT and 40 NT) will accumulate in the reaction mix. If the second step is faster, then you will see few intermediates, only the 250 NT and 85 NT final products.

You will use acrylamide gel electrophoresis to analyze the kinetics of this RNA splicing system. By running aliquots of the reaction mixture at various times during its progress, you should be able to inspect the gel and determine the relative concentrations of the different fragments produced during the reaction. This qualitative inspection allows one to determine which step is faster.

The laboratory experiment in this exercise is based on research done in the laboratory of Thomas Cech to characterize a self-splicing group I intron from the cyanobacterium *Anabaena* (Zaug et al. 1993). This experiment gives students the opportunity to carry out two important reactions that involve RNA: transcription and catalytic self-splicing. Students follow the time course of the two-step splicing reaction and determine the rate-limiting step. This experiment has been used successfully in the undergraduate biochemistry course at the University of Colorado, Boulder.

REQUIREMENTS

To perform this experiment successfully, you should have access to a laboratory equipped for work in basic molecular biology as well as the capability to run polyacrylamide gels. In addition, you should have experience with basic nucleic acid techniques. Some of the reagents are specialized, and unless you are in a research setting where they are already "on the shelf," it will be less time consuming, although more expensive, to use commercially available kits for steps such as plasmid isolation, *in vitro* transcription, and RNA isolation. Kits that have been used successfully with this experiment at the University of Colorado are indicated in the Reagent List. The experiment requires two lab periods (approximately six hours in all) and approximately two days of advance preparation work by you.

Background and Experimental Procedure

DAY 1, TRANSCRIPTION REACTION AND RNA ISOLATION

In the first part of the experiment, you will transcribe *in vitro* a linear version of the pAtRNA-1 plasmid. This recombinant plasmid encodes the catalytic intron sequence of interest and was linearized by cutting the plasmid with the *EarI* restriction enzyme. You will use this plasmid DNA as the template in the transcription reaction. To prevent the intron from self-splicing prematurely as it is being transcribed (rather than in the second part of this experiment), you will carry out the reaction under conditions of high nucleoside triphosphate (NTP) concentration. The reason is that self-splicing requires binding of a guanosine molecule to the folded precursor RNA (analogous to a substrate binding to an enzyme). The binding of guanosine, in turn, is promoted by Mg^{+2}, and so the negatively charged NTPs are present in the reaction mix to sequester this positively charged magnesium, thereby limiting the guanosine's ability to bind to the intron and stimulate its catalytic action.

After the transcription reaction, you will treat the mix with DNase (RNase-free) to remove all the high-molecular-weight plasmid DNA. You will then isolate the primary RNA transcript from any contaminating DNA fragments by binding the RNA to a silica-gel matrix, from which it will be eluted in purified form with water. You will perform both transcription and RNA isolation on the first day of the experiment.

DAY 2, SELF-SPLICING REACTION AND GEL ELECTROPHORESIS
In this part of the experiment, you will add Mg^{+2} to the purified RNA solution to allow it to self-splice. As self-splicing proceeds in the test tube, you will remove aliquots of the reaction at various times and quench them. You will then determine the size of the RNA fragments present at each time point by assaying them electrophoretically on an acrylamide gel. The gel separates molecules according to size, with smaller molecules moving farther. By visualizing the RNA fragments on the gel and comparing their migration distances with those of a set of standard RNAs of known size, you can assign sizes to the RNA fragments. You thus have an electrophoretic assay for the splicing reaction. You will relate the observed electrophoretic pattern of RNA fragments to one of two possible mechanisms for the two-step self-splicing reaction: first step faster or second step faster.

Before the class period in which students run the transcription reaction, you will need to do the following:

1. Prepare the necessary reagents (see Reagent List).

2. Grow the *E. coli* bacterial stock into which the *Anabaena* $tRNA^{leu}$ gene has been cloned.

3. Isolate from the bacteria the high-copy number plasmid DNA containing the insert.

4. Quantitate the plasmid DNA and treat it with *EarI* restriction enzyme to linearize the transcription template.

5. Grow the frozen stock cells overnight and isolate the plasmid DNA as follows:

   - Mix the stock cells in 10 ml of LB broth containing 50 µg/ml of ampicillin.
   - Dilute this suspension 100X into fresh LB plus ampicillin and grow overnight at 37°C.
   - Pellet cells and isolate plasmid DNA according to the instructions provided with the QIAGEN Plasmid Mega Kit (see Reagent List).

6. Quantitate the amount of DNA obtained by measuring its absorbance at 260 nm in a quartz cuvette and applying the following formula: [DNA] µg/µl = A_{260} x .050 x dilution factor. (Dilute the DNA in the cuvette as required to obtain an absorbance between 0.1 and 1.0.)

7. Treat the isolated plasmid with *EarI* restriction enzyme to linearize the template DNA for transcription by setting up a reaction tube containing the following reagents:

DNA	10X Restriction Buffer	*EarI*	H_2O
approx. 30 µg	50 µl	10 µl	to 500 µl

- Incubate the reaction overnight at 37°C.
- To ascertain that cutting has occurred, check aliquots of the cut and uncut DNA by running them on a 3% agarose gel.

8. Divide the remainder of the restriction reaction mix equally into 2 Eppendorff tubes. Precipitate and wash the DNA in each tube as follows:

- Add 1/3 volume of 7.5 M ammonium acetate.
- Fill both tubes with cold 95% ethanol; mix well.
- Allow DNA to precipitate overnight at -20°C.
- Microfuge tubes and wash the DNA with 70% ethanol.
- Air dry for 15 minutes and resuspend DNA in 10 µl of TE buffer or distilled water.
- Pool the contents of the two tubes and quantitate an aliquot by spectrophotometry as described above to ensure a concentration of approximately 2 µg/µl. Store the DNA at -20°C until needed for the transcription reaction.

Procedure

DAY 1, TRANSCRIPTION REACTION AND RNA ISOLATION

Your instructor will provide you with *EarI*-cut plasmid DNA (approximately 2 µg/µl concentration) as well as the reagents necessary for *in vitro* transcription. Use sterile technique in this experiment.

CAUTION: Wear gloves at all times.

Transcription Reaction

1. **Add the indicated number of microliters* of the following reagents to a sterile Eppendorff tube.**

DNA	Buffer**	H_2O	rNTP Mix***	RNase Inhibitor	T7 RNA Polymerase
20	5	48	10	2	5

* These volumes assume that you are not using a commercial transcription kit.
** 1 M Tris, pH 8.0
*** Prepare ahead by mixing equal volumes of the four 100 mM rNTP stock solutions. This mix provides a 4.0 mM final concentration of each rNTP in the reaction tube (16 mM in total rNTP).

2. **Mix the reagents with a pipette tip and incubate the tube at 40°C for 2 hours.**

3. **Following the first incubation, add 1µl of RNase-free DNase to the tube. Mix and incubate for an additional 30 minutes at 40°C.**

The reagent volumes indicated for the transcription reaction assume that you are using reagents you have prepared according to the specifications in the Reagent List. Although commercial kits for *in vitro* transcription are available from several

sources, the Mg^{2+} levels in the commercial transcription buffers are typically so high (60 mM in 10X buffer) that the intron will prematurely self-splice during transcription. If you elect to use a commercial kit, you should increase the rNTP concentrations in the reaction sufficiently to sequester this magnesium. To do this, you can add to the reaction tube a greater volume of the rNTP mix, while adding correspondingly less H_2O to keep the total reaction volume at 50 µl. In this case, you should check the reaction ahead of time.

Because the amount of template DNA is limited, have one group of students run the transcription reaction for the entire class. We suggest you run a control reaction to ensure that transcription has taken place. Materials to run such a control are provided in transcription kits. You might assign another group of students to run the control transcription reaction in place of the unknown. The control is set up in the same manner, except that 10 µl of a 250 µg/ml solution of *pSPT18-neo* control DNA is used in place of the *Anabaena* plasmid fragment.The control will give a 1,035 base transcript.

You also need to transcribe a set of standard RNAs of known size, so that the intron fragments can be sized on the gel. Use standard RNAs ranging in size from 500 to 100 NT, which are available commercially (see Reagent List). A third group of students can transcribe the RNA standards for the experiment in place of the experimental sample.

Isolation of RNA

After completing the transcription reaction, you will isolate the transcribed RNA for use later in the self-splicing reaction. You will use a commercial RNA isolation kit that employs a silica-gel column and provides all the necessary reagents.

1. **Add 50 µl of sterile water to the transcription mix tube.**
2. **Add 350 µl of RLT solution and 250 µl of 95% ethanol to the tube and mix by pipetting up and down.**
3. **Load the entire volume on the RNA isolation column provided.**
4. **Microfuge for 15 seconds.**
5. **Wash the column with 500 µl of RPE and collect the filtrate in a new Eppendorff tube. Watch the volume of wash solution accumulating in the collecting tube and dump it out when the fluid level in the RNA column gets close to the bottom of the column. You need to do this periodically, as the combined volume of the washes exceeds the volume of the collecting tube. Wash once more with 500 µl RPE. When the column has drained, microfuge it for 2 minutes to remove residual wash solution.**
6. **Get a sterile Eppendorff and elute the bound RNA into it with 50 µl of DEPC-treated water. Store the RNA solution in the -70°C freezer until it is needed for the self-splicing reaction.**

DAY 2, SELF-SPLICING REACTION AND GEL ELECTROPHORESIS

Now you will analyze the time course of the self-splicing reaction. You will initiate the splicing reaction by adding guanosine and a buffer containing Mg^{+2} to the purified catalytic RNA. As the reaction progresses, you will remove small aliquots at various times and dilute them into solution that stops the reaction of that sample. You then will run samples stopped at various times during the reaction on a polyacrylamide gel containing urea. Urea is a denaturant that eliminates RNA's three-dimensional structure. With no secondary or tertiary structure to complicate matters, the RNA fragments separate simply according to their molecular size. Based on the expected sizes of each fragment, you should see the following bands in an ideal separation:

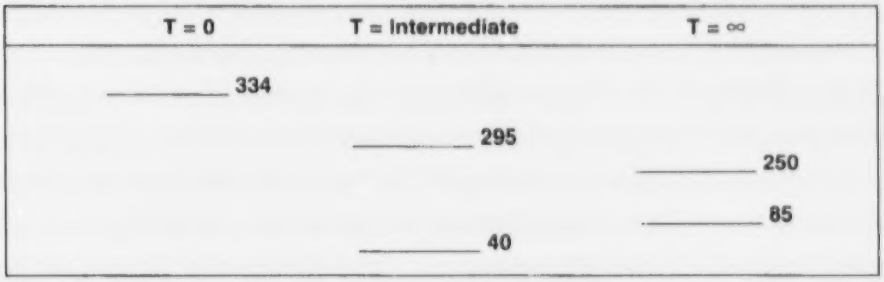


Figure 2.4 Distribution of bands expected on the acrylamide gel.

Splicing Reaction

1. **Set up four Eppendorff tubes, each containing 10 µl of stop solution. Label the tubes as follows: T = 0, T = 30, T = 60, and T = 90.**

2. **Set up a reaction tube containing 44.5 µl of the isolated RNA and 5 µl 10X HEPES buffer (with Mg^{++}). Mix and incubate at 50°C for 10 minutes. (Can you think of a reason for this preincubation?)**

3. **Begin timing as soon as you transfer the reaction tube to a water bath at 32°C. At the 1.5 minute mark, remove a 10 µl aliquot from the reaction tube and add it to the T = 0 stop solution tube.**

4. **At the 2.0 minute mark, add 5 µl 10X guanosine to the reaction tube and mix rapidly at 32°C. Every 30 seconds, transfer a 10 µl aliquot of the reaction to the appropriately labeled stop solution tube.**

Run the time-point samples on the polyacrylamide gel as described in the following section.

Polyacrylamide Gel Electrophoresis

Your instructor will demonstrate how to assemble the gel apparatus that you will use. You may use a prepoured gel, or you may pour your own gel. To pour a 9-by-11-inch gel, you will need to prepare about 120 ml of acrylamide solution.

1. **Dissolve 7.2 gm of acrylamide powder in 120 ml of 1X TBE buffer containing 8 M urea.**

CAUTION: Wear a mask and gloves when handling acrylamide powder.

2. **Add 1% by volume of freshly prepared 10% ammonium persulfate and 0.1% by volume of TEMED to the acrylamide-urea solution, mix, and carefully pour the solution between the glass plates. Insert the comb. Let the gel polymerize for at least 2 hours.**

3. **After polymerization, fix the gel plates into the electrophoresis unit, cover with an electrode buffer, and remove the comb. Immediately wash out the wells with a syringe containing electrode buffer. This creates good wells.**

   Prerun the gel with no samples for 30 minutes at 30 W, washing out the wells again after the prerun.

   Load the samples. Electrophorese at 30 W, 1,500 V until the slowest moving band, the xylene cyanol, is about 1 to 2 inches from the bottom of the gel.

4. **Turn off the power supply, then remove the gel and stain for 30 minutes in ethidium bromide. Photograph the gel.**

CAUTION: Observe appropriate precautions around electrophoresis equipment and when handling ethidium bromide.

Use either small-format, prepoured gels or larger format, 9-by-11-inch poured gels. To ensure good resolution of the RNA bands, let the gel run long enough for the xylene cyanol band to approach the bottom of the gel.

To save class time, you may want to pour the acrylamide gel(s) ahead of time. Small-format, precast, commercial gels may work but should be tested in advance to ensure adequate resolution of bands. An alternative to using ethidium bromide for staining is Stains-All (available from Sigma Chemical). Methylene blue may not provide sufficient sensitivity.

Challenge Questions

1. **Did transcription occur? How do you know?**

2. **Did splicing occur? How do you know?**

3. **Explain the gel data in your own words by drawing a diagram of the gel and describing the nature of the bands observed, such as "This band (indicate band with arrow) represents . . . "**

4. **Which step of the splicing reaction is faster?**

Additional Information

Zaug, A.J., McEvoy, M.M., & Cech, T.R. (1993). Self-splicing group I intron from *Anabaena* pre-tRNA: Requirements for base-pairing of the exons in the anticodon stem. *Biochemistry, 32:* 7946–7953.

REAGENT LIST

Instructors should prepare the following reagents as indicated:

a. ***EarI*-cut pAtRNA-1** (concentration approximately 2 μg/μl) - Prepare as described in Background and Experimental Procedure.

b. **Sterile 10X transcription buffer** - 1 M TrisHCl, pH 8.0. This buffer is formulated to be very low in magnesium. Remember, if you use a commercial transcription kit, the concentration of magnesium in the commercial buffer will be too high. In this case, you must increase the amount of rNTPs in the reaction mix as described under Day 1, Transcription Reaction, Step 3.

c. **Sterile 0.1 M $MgCl_2$** - autoclaved

d. **Sterile 10X nucleoside triphosphate solutions** - Four filter-sterilized aqueous solutions, each containing 100 mM of rCTP, rUTP, rATP, or rGTP (available as a set from Boehringer Mannheim, catalog # 1277057). Before running the transcription reaction, prepare a working rNTP mix by adding an equal volume of each rNTP stock to an Eppendorff tube. Remember, even if you use a commercial transcription kit, you need to prepare this rNTP mix to supplement the kit buffer, as the concentration of the rNTPs in the kit buffer is too low.

e. **Sterile 5 M NaCl** - autoclaved

f. **Sterile 10X HEPES buffer** - autoclaved 250 mM HEPES buffer, pH 7.5, containing 100 mM $MgCl_2$

g. **Sterile 10X guanosine** - 250 μM guanosine in 1X HEPES buffer. Guanosine crystal is available from several suppliers, such as Fisher or Sigma Chemical.

h. **1X Tris-borate-EDTA buffer (TBE)** - 0.1 M Tris base containing 0.083 M boric acid and 1 mM EDTA

i. **Stop solution** - 0.1X TBE containing 30 mM EDTA, 10 M urea, 0.01% bromophenol blue, and 0.025% xylene cyanol. Bromophenol blue-xylene cyanol dye powder are available from Sigma Chemical.

j. **Polyacrylamide gel** - 6% polyacrylamide plus 8 M urea in 1X TBE

k. **DNase I (RNase-Free)** - Available separately from Ambion, Inc., catalog # 2222 (2 U/μl). Use at the concentration supplied. Ambion, Inc., 2130 Woodward Street, Suite 200, Austin, TX 78744 (512) 445-6979. This reagent is provided in transcription kits.

l. **T7 RNA polymerase** - Available separately from Ambion, catalog # 2084 (T7 RNA polymerase—Cloned, High Con—200 U/μl, 30,000 U) or provided in transcription kits.

m. **RNase inhibitor** - Available from Ambion (Rnase Inhibitor-Cloned, catalog # 2682, 2,500 U at 40 U/μl). Also available from Promega as Rnasin®, catalog # N2511 (2,500 U at 20 to 40 U/μl).

In addition to the reagents listed above, you will need the following specialized materials:

Commercial Kits and Reagents

n. **Plasmid DNA isolation** - QIAGEN Plasmid Mega Kit, catalog # 12181 (2.5 mg column capacity). Qiagen, Inc., 28159 Avenue Stanford, Valencia, CA 91355-1106 (800) 718-2056.

o. **RNA isolation kit** - QIAGEN Rneasy Mini Kit, catalog # 74103 (20-reaction kit).

p. **RNA size standards** - (500 to 100 nucleotides) Ambion Century™ Marker Template, catalog # 7780 (5μg).

q. ***Earl* restriction enzyme** - New England Biolabs (10 U/μl), includes 10X buffer.

r. **(Optional) transcription kit** - Boehringer Mannheim Kit SP6/T7, catalog # 999644 (40 reactions) or Promega Riboprobe® Combination System SP6/T7, catalog # P1460 (25 reactions). Remember to compensate for the high magnesium levels in the kit buffers by adding additional rNTPs to the reaction mix. (See discussion under Day 1, Transcription Reaction, Step 3.)

Biological Material

s. See Zaug et al. (1993) in *References and Related Literature*.

FOR YOUR INFORMATION

The Origin of Life: New Answers for an Ancient Question

The origin of life has always been a fascinating and vexing question. Throughout the history of humankind, this mystery has been the subject of passionate debate, and every age has come up with answers that reflect the religious, philosophical, and scientific beliefs of the time. During the last 400 years, science and technology have been extraordinarily successful in allowing us to understand and manipulate the physical world. More recently, the biological sciences have provided an understanding of and ability to manipulate living systems that is unprecedented in history. In light of such modern advances, it is only natural that we ask the ancient question anew: Where did we, and all life, come from?

For some, the answer is simple, satisfying, and ancient: God or a deity made all life as we know it.

Although the existence of a deity is not a topic that can be addressed through scientific inquiry, modern explanations of life's origin need not be based on received faith or unfounded opinions. We can seek a valid naturalistic explanation for the mechanism of life's origin, an explanation that is consistent with experimental evidence. New data from ongoing scientific research is providing the means to construct informed and testable hypotheses of our origins.

63

FOR YOUR INFORMATION (continued)

Of the recent data bearing on the origin of life, perhaps the most revolutionary was the discovery in 1981 that RNA can be catalytic. This discovery was significant to the question of life's origin because it showed for the first time that a molecule, RNA, exists that has both of the key properties essential to the beginnings of a living, evolving system: the ability to *encode information* and the *catalytic ability* to conceivably replicate itself and other molecules.

The discovery of catalytic RNA resolved a chicken-egg paradox about whether nucleic acids or functional protein catalysts evolved first, which had long stymied progress in finding a viable scientific explanation for the origin of life. If the nucleic acid RNA can function as a catalyst, then it could in principle have functioned early in evolution as both *genome* (where genetic information is stored as the linear sequence of nucleotide bases) and *replicase* (the catalyst that replicates the genome). No DNA or proteins would be required for an RNA molecule to duplicate itself in this hypothetical RNA world. This RNA molecule could then evolve through natural selection into even more complex RNAs possessing additional coding and catalytic functions.

Eventually in the course of evolution, DNA took over as the primary genetic material, while proteins took over the vast majority of catalytic functions. All organisms today (except the RNA viruses) have a genome of DNA. One reason that DNA may have taken over from—that is, outcompeted—RNA is that it is more chemically stable than RNA. RNA slowly breaks down into monomer units even at room temperature, whereas DNA can persist for very long periods, long enough to enable biologists to recover DNA from a 40,000-year-old wooly mammoth (but not from a 70-million-year-old velociraptor). Proteins later emerged as the primary biological catalysts because of the chemical versatility of their 20 amino acid building blocks and their ability to assume a seemingly endless number of potential shapes.

The progenitor RNA almost certainly had a crucial ability conferred by its coding and catalytic properties: the ability to reproduce itself. This property of molecular replication gets to the very heart of what it means to be "living." The distinguished geneticist and evolutionary biologist H.J. Muller, responding to the question of what it means to be a living system, has said, "I think the most fundamental property distinguishing a living thing . . . is its ability to form copies of itself. We call this 'reproduction' . . . " To further clarify where the origin of life resides, Muller went on to say, "I should draw the line where the Darwinian process of natural selection begins to come in, and that is at the appearance of (molecular) replication of a self-copying kind—that is, the replication of mutations."

Our working hypothesis for the origin of life, the RNA world hypothesis, is that the first "living" and evolving entity is likely to have been a self-replicating RNA molecule. In the following activity, you will have a chance to explore RNA replication further.

Admittedly, naked replicating molecules are a long way from the complexity of even the simplest known cell. But laboratory approaches like those guided by the RNA world hypothesis provide for the first time what has been lacking in origin of life research: an experimentally testable starting point for the first steps in the evolutionary process.

Activity 3
RNA and Evolution

Activity at a Glance

Exercise 3.1: Students model nucleotide base interactions and discover non-standard base pairings.

Exercise 3.2: Students learn about the strategy of templating, and they model RNA replication in an attempt to discover a hypothetical pathway for RNA self-replication.

Exercise 3.3: Students learn about a classic study of RNA evolution *in vitro*, either by going online to simulate the experiment interactively or by completing a paper and pencil activity.

Exercise 3.4: Students apply their understanding of templating to interpret a new experimental system for continuous *in vitro* replication of RNA.

Exercise 3.5: Students have the opportunity to conduct a continuous *in vitro* replication experiment in the laboratory.

Concepts

- The origin of life is an issue that can be addressed using scientifically valid evidence to support an explanation.
- The wide range of modern RNA function makes it a prime candidate for early life in an RNA world.
- Life and its evolution require a means to replicate genetic information.
- Molecular replication, selection, and evolution can be demonstrated *in vitro*.
- In nature, evolution does not proceed by design.

Focus

Students consider the essentials of life, explore the nature and significance of genetic replication, and, most importantly, work with team members to construct a well-reasoned explanation for the origin of life based on scientifically valid evidence.

BEST COPY AVAILABLE

Connections to Lecture Topics

Use Activity 3 in conjunction with a study of nucleic acid replication or as part of the study of evolution and natural selection.

Estimated Time

Approximately 2 hours in class and 2 hours of laboratory time

Materials Preparation

For each team of four students, provide
- nucleotide base cutouts (see *Copymasters*)
- millimeter ruler
- opaque tape
- access to the World Wide Web (optional)

See annotation comments and reagent lists in Exercise 3.5 for laboratory materials.

ANNOTATED STUDENT ACTIVITY

Introduction

CRICK'S CENTRAL DOGMA

In 1968, Francis Crick neatly summarized what was then our understanding of the flow of genetic information in living systems. He proposed what is still referred to as the central dogma of molecular biology. The dogma asserts that in the biological world, the translation of genetic information is fundamentally a one-way process: *from* nucleic acids, where information is stored as the "language" of specific sequences of nucleotide bases, *into* the "language" of proteins, whose functional amino acids and diverse shapes allow them to express (do something with) the information in the cell. Crick's view of the central dogma is summarized in Figure 3.1.

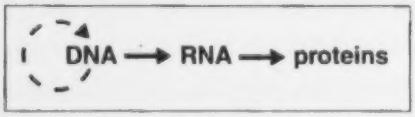


Figure 3.1 Crick's central dogma. The curved arrow indicates DNA's well-known ability to be copied during replication.

The choice of the term *dogma* to describe what is essentially a scientific hypothesis was unfortunate. A dogma, because it implies an unchanging and rigidly held belief, is just the opposite of a scientific hypothesis. A hypothesis is a working proposal that is used to guide future experiments, is held tentatively, and is subject to modification or rejection at any time as new evidence dictates. True to its real nature as a scientific hypothesis, Crick's original proposal has been modified significantly as our understanding of genetic information has grown during the last 30 years.

Crick's main assertion, that genetic information flows from nucleic acids to proteins, and not in the reverse direction, has stood the test of time. No instance of reverse translation from proteins to nucleic acids has ever been observed. But today we know that Crick's original scheme of information flow was incomplete. Newer research reveals that genetic information can flow back and forth *between* the nucleic acids DNA and RNA. The well-known path of transcription of DNA into RNAs has been widened into a two-way street as examples of transcription of RNAs into DNA have been discovered. Some examples of this reverse transcription are the copying of viral genomic RNA into DNA, with the DNA's subsequent insertion into the host cell genome (for example, HIV and

other *retro*viruses). Another example is the occasional copying of cellular mRNAs back into DNA form, with their insertion back into the genome (as pseudogenes). The 1970 discovery of the enzyme reverse transcriptase, and the two-way informational "crosstalk" between DNA and RNA that it makes possible, was so significant that it earned David Baltimore and Howard Temin a Nobel Prize and required a modification of the original scheme for genetic information flow (Figure 3.2).

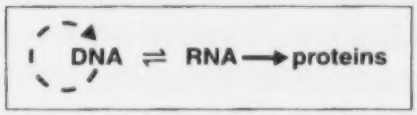


Figure 3.2 Baltimore and Temin's modified scheme for the flow of genetic information.

THE CENTRAL DOGMA SPAWNS THE CHICKEN AND THE EGG

A good hypothesis typically yields new questions to be answered; the central dogma was no exception. Biologists interested in the origin and evolution of genetic information mechanisms have long faced a dilemma in trying to figure out which key cellular process, *DNA synthesis* (replication) or *protein synthesis* (translation), was the first to evolve. While trying to answer this question in light of the known biochemistry of contemporary cells and the information flow shown in Figure 3.1, a classic chicken and egg paradox presented itself: If DNA can only be assembled and copied with the aid of protein enzymes, and protein enzymes can only be encoded by a pre-existing DNA, then logically neither one could have arisen first without the other also being present.

THE CHICKEN AND THE EGG MEET THE RNA WORLD

This paradox was apparent to Crick when he proposed his original scheme of information flow. Indeed, he was sufficiently troubled by it that he was one of the first to propose a theoretical solution: Perhaps *RNA* was the first genetic information molecule, emerging before either DNA or protein, and later giving rise to both. This novel proposal remained an interesting speculation for 15 years because, during this time, biologists had no experimental evidence that RNA possessed the biochemical versatility necessary to fulfill the role of progenitor molecule. All of that changed in the early 1980s, when Thomas Cech and Sidney Altman independently discovered the catalytic ability of RNA.

A world of possibilities concerning the origin of life and early evolution opened to biologists once they realized that RNA can potentially fill *both* key roles required of an evolutionary progenitor molecule: encode genetic information and act as a functional catalyst capable of synthesizing other molecules, perhaps DNA and proteins. This new vista is termed the *RNA world hypothesis*, which proposes that RNA originated first and functioned in the earliest stages of molecular evolution as both the encoder of genetic information and the catalytic worker molecule. If correct, this hypothesis eliminates the chicken-egg dilemma (DNA first or protein first) by proposing that RNA came first. But just what functions did these versatile primordial RNAs have that initiated biological evolution at the molecular level? In this activity, you will examine some of the key functions of a primordial RNA.

The RNA world hypothesis expands once again our working scheme of biological information flow and puts it into an evolutionary, historical perspective (Figure 3.3).

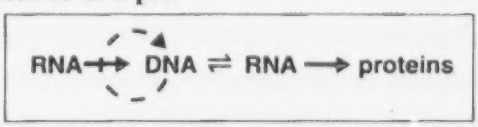


Figure 3.3 The RNA world hypothesis.

Challenge Questions

Read the Introduction, then answer the following questions:

1. **Discuss some attributes common to all life forms and generate a brief list of those attributes that you feel are the most essential. Why have you selected these attributes?**

   Some of the most essential attributes common to all life forms include the ability to transform energy, the ability to encode and use chemical information, and the ability to reproduce.

2. **If the RNA world hypothesis is correct, and RNA was the first heritable biomolecule, then RNA must have used both its information coding and catalytic abilities to accomplish a task essential for its perpetuation. Discuss what this task would be. Is this task included on your list of life's essential attributes?**

   If the RNA world hypothesis is correct, and RNA was the first heritable biomolecule, then RNA must have used both its information coding and catalytic abilities to replicate itself.

3. **Modify the information flow scheme shown in Figure 3.3 to reflect your answer to Question 2.**

   In Figure 3.3, students should add a curved arrow from RNA back to itself, in order to reflect the self-replication of RNA.

Exercise 3.1: Nucleotide Base Complementarity

In this exercise, you will explore the concept of molecular self-replication. In particular, you will attempt to determine how, very early in the course of biological evolution, a hypothetical RNA molecule might have made copies of itself without the aid of protein enzymes. This property of unaided self-replication is not the rule in living systems today: All DNA and RNA that we know of is replicated by protein enzymes (polymerases or replicases) that assemble the monomer building blocks of these nucleic acids into new copies. Why do we think that RNA molecules might once have been, and may still be, capable of copying themselves?

The rationale stems from the RNA world hypothesis and from the observation that certain catalytic RNA molecules are known to promote the ligation, or joining together, of nucleotide building blocks, while using their own sequence as a template. If a primordial RNA was able to combine information coding and catalytic abilities in this way, it may have been able to make complete copies of itself. This self-replication would have started evolution as the RNA proliferated and inevitably produced variants of itself. Different variants would replicate more or less rapidly, would "compete" with one another for raw materials, and some would eventually assume new functions, such as the synthesis of DNA and proteins.

We are all aware of the universal process of biological reproduction, especially when it results in the creation of a new individual—a plant giving rise

to a new plant, or parents giving birth to a child. As biologists, we need to be aware that *all reproduction at the organismal level begins with molecular reproduction—replication—at the level of the genetic molecules DNA and RNA.* At life's most basic level, then, molecular replication is essential.

What are the properties of DNA and RNA molecules that make their replication possible? Quite simply, they are *complementary interaction* and *templating*. The term complementary interaction is used here in the chemical sense: Two molecules interact in a complementary way if their shapes "fit" one another (like a hand in a glove) and they are able to form stable bonding interactions as a result. You know from Activity 1 that in the case of the nucleic acids it is the monomer nucleotide bases A, U (T in DNA), G, and C that engage in such complementary interactions. The interactions follow defined pairing rules, based on the ability of certain base pairs to form stabilizing hydrogen bonds. All known biological systems use this principle of complementarity to replicate their genetic material.

Activity 1 acknowledged the complementary interactions between nucleotide bases and was concerned with them only as determinants of the three-dimensional folding of RNA molecules. Here you will examine the physical interactions of complementary bases more closely and consider the importance of these interactions for the replication of nucleic acids.

Materials

- nucleotide base cutouts
- millimeter ruler
- opaque tape

Procedure

1. Arrange paper cutout models of several different monomer nucleotide bases randomly on the desk. Some of the models represent pyrimidine bases, made up of a single ring, and others are purine bases, containing two rings.

2. Group the bases into stable hydrogen-bonded pairs. The dashed lines (- - - -) on the models represent hydrogen bonds. Not all interactions among the bases are stable, but stable interactions will meet the following criteria:

   - An H on one molecule will form a hydrogen bond (- - - -) with an O or an N on the partner molecule.
   - At least two hydrogen bonds between a pair are needed for a stable interaction.

3. List below the stable pairs and the number of hydrogen bonds in each. Also measure the overall diameter of each pair.

Base Pair	Hydrogen Bonds (#)	Diameter (mm)

4. **Cover the dashed lines (- - - -) in the uracil (U) with opaque tape and see if you can find another partner for it. Record your findings above.**

Challenge Questions

1. **How do the stable pairs that you found relate to the purine and pyrimidine base classes?**

   Each stable pair is made up of a purine plus a pyrimidine base.

2. **Why are only some base pairings stable? Speculate about the relative stability of pairs with two bonds versus those with three.**

   Only certain pairs meet the criteria for the number and composition of H bonds.

3. **What pairing(s) did you find for adenine (A)? Which is relevant to DNA? to RNA?**

   A-T (DNA) and A-U (RNA).

4. **What two pairings did you find for uracil (U)? Which of these is a "standard" Watson-Crick pairing? The alternate pairing for U is referred to as a "wobble" base pair. Wobble pairs and a few other nonstandard pairings increase the number and variety of base associations. Predict the general effect of these diverse pairings on the variety of shapes that RNA molecules can potentially assume.**

   A-U (standard Watson-Crick pair) and G-U (wobble pair). Nonstandard pairings increase the potential variety of RNA shapes.

5. **How do the relative diameters of stable pairs compare? Assume that the model bases have the same relative geometric proportions as the actual bases; what implication does of your observation on base-pair diameters have for the overall diameter of a double-stranded nucleic acid? How might a G-G mispairing influence the diameter of the double-stranded molecule?**

   Diameters are roughly the same for each pair. The implication is that a double-stranded nucleic acid has a uniform overall diameter. Mispairings such as G-G or A-A create a local "bulge" in the molecule. These bulges are referred to as bulge loops, and they have a destabilizing influence on the double-stranded region.

Here students model base-pairing interactions using paper cutouts. You will initially give cutouts of the four standard RNA bases adenine (A), uracil (U), guanine (G), and cytosine (C) (see *Copymasters*). As they attempt to pair the cutouts, students should discover the standard Watson-Crick RNA base pairs (A-U and G-C). Students may also discover "wobble" pairings between G-U. Next give students a thymine model and ask them to compare the A-U and A-T base pairs to understand how RNA and DNA are essentially identical.

Students next measure the diameter of the different base pairs to discover their essential constancy.

74A

As an extension, you could give students a cutout of the base inosine (a modified purine common in RNA) and allow them to discover that it has multiple pairing partners (C, U, or A). Encourage students to speculate on the implications of such alternative pairings for the diversity of RNA shape.

As an introduction to the phenomenon of replication by complementarity, give students a cutout of a linear RNA oligomer with a fixed sequence of the standard bases and ask them to align the monomers along the oligomer template.

As an extension to illustrate the generality of the phenomenon of replication by molecular complementarity, give students cutouts of the small synthetic self-replicating catalysts developed by Julius Rebek. See Rebek, J. (1994). Synthetic self-replicating molecules. *Scientific American*, 271(1): 48–55.

Next you will use the strategy of molecular templating to explore how complementary interactions of bases play a role in the replication of nucleic acid polymers.

You may already be familiar with using a template to make a replica. A sculptor, for example, can employ a template to produce a statue (Figure 3.4): In Step 1, a hollow mold (the template) is filled with liquid material that hardens. Separation of the template from the hardened material yields a product that is *complementary* to the template, in the sense that the product is solid where the template is hollow, and the surface contours of the product protrude where those of the template recede.

Challenge Questions

Refer to Step 2 in Figure 3.4. Assume that the sculptor uses the statue itself, the complementary product, as a new template and pours liquid material over its surface. After the material hardens, the sculptor separates the products.

1. **How does the hollow end-product of this second step compare with the objects in Step 1?**

   It is a copy of the original template and complementary to the statue.

2. **What is the overall result of this process?**

   Copying, or replication, of the original template.

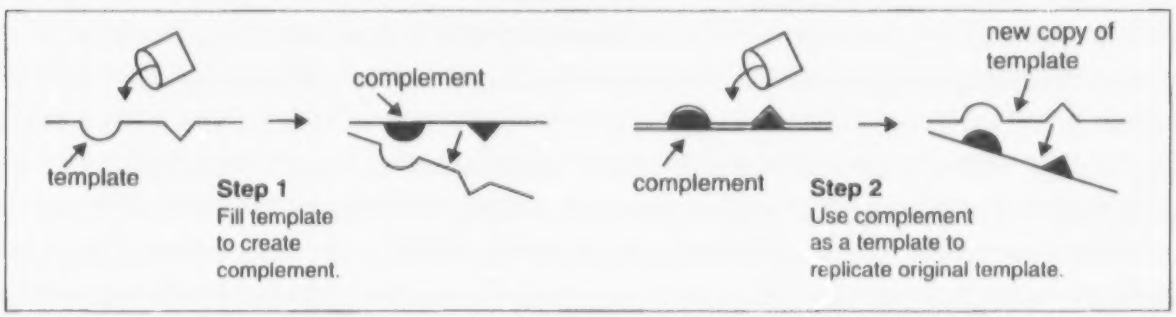


Figure 3.4 Using a template to make an exact copy of an original.

Exercises 3.2 and 3.3 are available as interactive activities on the World Wide Web at The Biology Place, a subscription site hosted by Peregrine Publishers (www.biology.com/). If you do not have access to this site, you can have the students conduct the exercises with the paper and pencil versions presented below.

Exercise 3.2a: Discovering a Pathway to Self-Replication

Because our working hypothesis is that a self-replicating RNA is likely to have been the first "living" and evolving molecule, we must test this hypothesis by finding such a molecule in nature or creating one in the laboratory. As yet, no naturally occurring RNA able to fully replicate itself has been found in any organism. Such an RNA could still exist, perhaps in a hot spring or deep ocean vent, as the living fossil that would provide evidence of a past RNA world. Biologists are on the lookout, and time will tell. In the meantime, scientists are developing new experimental approaches that have real potential to discover just such a self-replicating RNA in the laboratory. You will explore this laboratory approach in a later exercise.

In this exercise, you will discover a sequence of events or a mechanism by which self-replication might take place.

Procedure

The following single-stranded RNA sequence represents a hypothetical ribozyme that can act as a generalized ligase able to couple nucleotide monomers and form a new single-stranded RNA molecule. The ribozyme can use either itself or other RNA molecules as a template to guide this assembly. Also shown are supplies of the four nucleotide building blocks, A, U, G, and C. Your goal is to devise a series of steps by which this template/ligase molecule can ultimately assemble an *exact copy* of itself.

1. **Apply base pairing and the template strategy while writing in the appropriate monomers next to the template. After completing this step, answer the Challenge Questions that follow.**

Template/Ligase Ribozyme	Monomer Pools	
5′		
C		
G	A A A	C C C
U	A A A	C C C
C		
G	G G G	U U U
G	G G G	U U U
A		
A		
A		
C		
G		
U		
A		
U		
C		
C		
3′		

75 A

Challenge Questions

1. **Which of the following terms applies to the new single-stranded molecule that you assembled?**

   a. exact copy
   b. complement

   b. Complement.

2. **Is the molecule that you assembled likely to be able to act as a template? Explain.**

   Yes. Any RNA sequence can be a potential template. What is needed is a catalytic ligase molecule that can use the template to construct a new molecule.

3. **Is the molecule that you assembled able to also act as a ligase catalyst? Explain.**

   Not necessarily. The complementary sequence of a ribozyme RNA is not necessarily also catalytic. It is important to realize that even though the template molecule in this activity is depicted as a rigid linear shape, real RNA molecules are very flexible and able to fold into a variety of three-dimensional shapes. For any given nucleotide sequence, the molecule's shape is determined by complementary pairing between the bases within the molecule itself (see if you can determine the shape that internal base pairing would confer on the template RNA in this activity).The catalytic activity of RNAs that possess it is due to their particular base sequence and resulting three-dimensional shape. Notice that the complementary molecule students made has a different primary sequence of bases than its template. As a result, the complement will fold into a different three-dimensional shape. This shape may or may not have any catalytic ability; in fact, it is unlikely to because catalytic RNAs are rather uncommon. And even if it were catalytic, the specific type of reaction catalyzed would likely be different from that of its complement.

Procedure (continued)

2. **Assume for the moment that the complementary molecule you constructed *is* a catalytic ligase, like the original template/ligase molecule. How might an exact copy of the original be generated?**

   This is the simplest but least likely case, in which the the complement now simply uses itself as a template to assemble its own complement. This product would be an exact duplicate of the original molecule. This has yet to be observed in the laboratory.

3. **Now assume a different scenario in which the complement that you initially assembled is *not* catalytic. Proceed again to produce the exact copy of the original template RNA. Hint: Consider where the ligase activity might come from for this next step. Keep in mind the properties of the original template/ligase molecule that you started with.**

In this case, the generalized catalytic activity must come from the original template/ligase. In principle, it could use its own complement as an external substrate, acting on it much the way a protein replicase would to produce a new polymer. This complement of a complement is the duplicate sought. Although this scenario is somewhat more probable than the previous one, it is still improbable. Nevertheless, it is in principle possible, and researchers are actively seeking an RNA with this ability.

Challenge Questions

1. **How did you solve the puzzle of how to carry out the second step of the self-replication process?**

   The starting RNA was able to act as both a template and a ligase catalyst. In the first step of the process, the RNA used *both* abilities together to assemble its complement: It used itself as a template against which to join bases complementary to its own. If students find it difficult to visualize the template/ligase doing this, ask them to imagine themselves placing complementary blue dots against each of the green buttons of their shirts.

   In the second step, the starting RNA molecule functioned only as a ligase, binding to its complement molecule and using it as a template. The result of this second round of copying was an RNA molecule identical to the starting sequence (the complement of a complement is a copy of the original). Now students can see why this copying process is referred to as *template-directed self-replication*.

2. **A two-step process may be the simplest path to self-replication, but it is unlikely to happen this way. More likely, the process is gradual, with several steps involved. Can you think of another route to self-replication?**

   It has been proposed that an alternating sequence of untemplated and templated steps gradually built up the first RNA able to act as both template and ligase. The untemplated steps would have consisted of the random joining of nucleotides, perhaps catalyzed by minerals. These random molecules could then have acted as templates for the mineral-catalyzed formation of complements. A catalytic RNA with ligase function might eventually have been produced in this way.

Additional Information

Orgel, L.E. (1994, October). The origin of life on the earth. *Scientific American*, *271*(4): 77–83.

Joyce, G.F., & Orgel, L.E. (1993). Prospects for understanding the RNA world. In Gesteland, R.F., & Atkins, J.F. (Eds.), *The RNA world* (pp. 1–25). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

FOR YOUR INFORMATION
Modern Evidence of an Ancient World

As we study the many functions of RNA in contemporary life, we see remnants of early informational and catalytic molecules that could have taken the first unaided steps in building the maze of tiny bridges that led from inanimate chemicals to living creatures. These versatile RNA molecules likely populated a primordial RNA world, in which the process of evolution first began.

The efficient genomes of certain contemporary viruses and even simpler viruslike agents demonstrate that RNA might easily have been a major genetic player in the saga of life. And today's remaining catalytic RNAs (ribozymes) may well be the 4-billion-year-old descendants of early catalysts that made possible key evolutionary processes, such as the replication of RNA itself. Evidence of RNA's early importance mounts as new contemporary RNAs and their diverse functions in the cell continue to be discovered. The familiar mRNAs, rRNAs, and tRNAs involved in gene expression and protein synthesis have been joined by several recently identified molecules, such as the small nuclear RNAs (snRNAs) that are involved in splicing and editing cellular mRNAs. New functions are even being recognized for some of the familiar RNAs. Transfer RNA, for example, has been found to participate in nucleic acid replication and in the synthesis of bacterial cell walls. Some ribonucleotides, the monomer building blocks of RNA, are key components of the coenzymes that assist protein catalysts. Modern coenzymes may be leftovers from an ancient RNA world in which primitive catalysis and metabolism were conducted solely by RNA molecules that eventually acquired the ability to synthesize more efficient proteins. These and many other examples show us that the range of functions for RNA is considerably wider than previously thought (Figure 3.5).

In our search for evidence of RNA's versatility and life's origins, we are not limited to identifying molecular functions that exist in modern cells or viruses. Research scientists now conduct laboratory experiments in which populations of RNA molecules are made to

- RNA stores information and performs catalysis *in vivo*; no other biomolecule has both properties.
- Nucleotide sequences of RNAs common to all organisms (for example, rRNAs) are highly conserved (similar) among the many different species studied, suggesting that RNA was a key molecule present early in evolution.
- RNA or ribonucleotides are involved in most critical cellular functions in all three domains of life:
  - Adenosine triphosphate (ATP) is a universal energy carrier.
  - Universal metabolic pathways employ adenine nucleotide coenzymes (NADH, NADPH, FAD, CoA).
  - Protein synthesis employs mRNAs, rRNAs, and tRNAs.
  - rRNA by itself can catalyze peptide bond formation.
  - DNA synthesis requires the prior conversion of ribonucleotides to their deoxy form.
  - The ribonucleotide uracil, found only in RNA, is the precursor for DNA's thymine.
  - RNA is the primer for DNA replication.
  - Ribonucleotide derivatives function as key signaling molecules in the cell (for example, cAMP, ATP).
- RNAs function as primers in DNA replication and reverse transcription of retroviral genomes.
- tRNA-like molecules are involved in nontranslational polymerizations (for example, cell wall synthesis, antibiotic synthesis).
- A tRNA-like molecule may have given rise to the RNA component of telomerase, the enzyme that maintains the ends of chromosomes.
- Enzymatic processing of mRNAs involves other small RNAs (for example, snRNAs, RNase P).
- Protein sorting into the endoplasmic reticulum of all eukaryotes involves RNA (for example, SRP-RNA).
- During polysaccharide synthesis, ribonucleotides activate and carry sugars.

Figure 3.5 Modern RNA functions that are consistent with an early RNA world.

FOR YOUR INFORMATION (continued)

undergo evolution in the test tube, producing molecules with entirely novel structures and functions. These *in vitro* selection experiments demonstrate that populations of RNA evolve according to known principles.

The evidence at hand from several different approaches raises our confidence about ancient scenarios that envision RNA as perhaps the first self-replicating molecule—a pioneer of life.

Although no RNA able to completely replicate itself has yet been found in nature, scientists have made progress toward discovering a self-replicator in the laboratory. These experiments are conducted in the test tube, at the molecular level, and apply the key elements of evolution: *variation*, *selection*, and *replication*. Scientists are learning much about the potential biochemical capabilities of nucleic acids and are developing evidence-based hypotheses about how the first evolving RNAs might have come to be. In Exercise 3.3, you will explore an early and pioneering laboratory approach to the study of molecular evolution.

Exercise 3.3 is available as an interactive World Wide Web-based exercise at The Biology Place, a subscription site hosted by Peregrine Publishers (www.biology.com/). The Web version allows students to simulate a classic experiment in molecular evolution and to generate data for analysis. If you do not have access to this site, your students can work with the paper and pencil version of the exercise that follows.

Exercise 3.3: Molecular Evolution in the Test Tube

In the mid-1960s, researcher Sol Spiegelman developed the first system for studying the replication and evolution of RNA molecules in the test tube (*in vitro*). As his starting material, Spiegelman chose the RNA molecule comprising the genome of bacteriophage Qβ, a virus that normally infects the bacterium *Escherichia coli* (*E. coli*). To be replicated in a natural infection cycle, the Qβ RNA must first get inside an intact *E. coli* cell. To accomplish this, three of the four genes encoded by the RNA's 4,000 nucleotides specify proteins that enable the RNA to enter the bacterial cell and its "progeny" RNAs to spread to new bacteria. The fourth gene encodes viral replicase, the protein enzyme that uses the viral RNA as a template on which to assemble monomers into new copies of the RNA. The replicase enzyme initiates copying of the RNA by first binding to a small subset of bases within it, called the *origin of replication*. These few bases are all that any Qβ RNA molecule needs to be copied by the replicase. Any molecule with an intact origin of replication sequence will be copied, and any molecule in which this sequence is either lost or significantly mutated will not be copied or will be copied at an altered rate.

The experimental system that Spiegelman employed was well suited to the study of molecular replication and evolution *in vitro*:

- The experimental system streamlined and simplified the viral RNA replication process, as compared with a natural *in vivo* infection. Specifically, the system eliminated the requirement for Qβ RNA to first get inside an intact bacterial cell. Spiegelman accomplished this by providing free in the reaction tube all the raw materials needed for RNA synthesis (that is, the A, U, G, and C building block nucleotides of RNA, plus some accessory bacterial proteins). These materials normally would only be available to Qβ once it is inside a host bacterial cell. Also provided in the tube was an ample supply of the viral replicase enzyme.
- The system also had built-in mutation features ensuring base sequence changes (molecular variation) in the population of "progeny." First, the replicase enzyme is a relatively sloppy worker, making one or two random nucleotide base changes (mutations) in each RNA copy that it produces. Second, the replicase occasionally produces randomly broken copies of the RNA. In a natural infection of *E. coli*, many of these defective, shorter-than-normal RNAs would be uncopyable or unable to spread the infection to other bacteria. Broken molecules in the test tube system, in contrast, were not at a disadvantage for either copying or perpetuation into the next "generation." They could still be copied, provided that *they retained intact the short origin of replication sequence recognized by the replicase enzyme.* And, as described above, their perpetuation was independent of their ability to infect cells in this test tube system.

The test tube system, then, neatly incorporated two of the three features essential for all evolution: replication and variation.

Spiegelman next provided the third essential feature of evolution, selection, to observe the evolution of the starting population of RNA molecules. To apply a *selective pressure* (a condition favoring some individuals and disfavoring others) on the population of molecules, Spiegelman limited the time available to complete replication, thus giving a selective advantage to those RNA molecules that could be copied quickly. In effect, this time limit transformed the copying process into a race between variant RNAs that arose at random during the course of the experiment. Speed of replication became a "phenotypic trait" of the molecules and a test of their molecular "fitness" in this test tube "environment."

Spiegelman and his associate started the experiment by adding Qβ RNA to a test tube containing replicase enzyme and nucleotide monomers. They allowed the replication reactions to proceed for just 15 minutes. They then transferred a random sample of progeny RNAs from the first tube into a second tube containing a fresh supply of nucleotides and replicase enzyme (but no RNA other than that was transferred). The replication process was again allowed to proceed for 15 minutes, and a sample from this second tube was transferred to a third tube of fresh raw materials. This serial-transfer process was repeated 73 times (Figure 3.6). During the reaction, the experimenters monitored the total amount of RNA that accumulated in each tube as well as the size and nucleotide base composition of each "generation" of RNA.

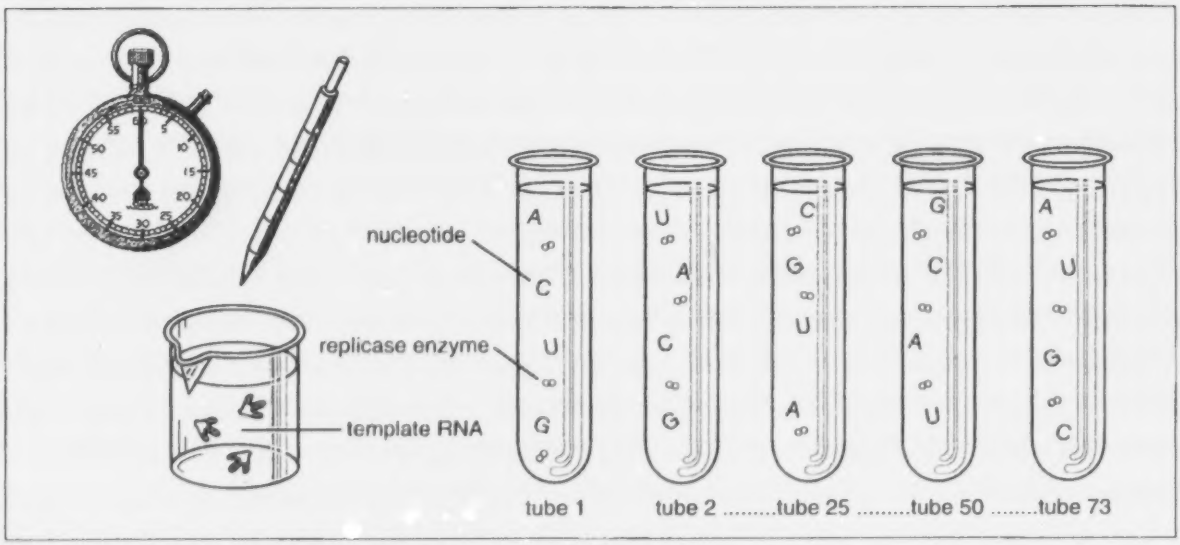


Figure 3.6 Experimental setup for *in vitro* replication of RNA.

Procedure

1. **Use what you now know about Spiegelman's experimental system and about evolution to formulate a hypothesis and make general predictions about how the starting RNA population was altered during the course of Spiegelman's experiment. Your hypothesis should make predictions about changes that might occur in three phenotypic traits of the RNA molecules:**

   - **speed of replication**
   - **length of the molecule**
   - **ability to infect *E. coli* bacteria**

Challenge Questions

Revise your initial hypothesis if necessary in light of the data in Figure 3.7 and answer the following questions:

1. **What can you conclude from Figure 3.7 about the total number of RNA molecules produced in each generation? What does this imply about the average replication speed of the molecules?**

   The total number of molecules replicated in each generation increased progressively. The average speed of replication per molecule increased dramatically. The molecules in the final population were being copied 15 times faster than the molecules in the starting population. Students should recall that Spiegelman's selection criterion was replication speed under the conditions of the experiment.

2. **The reason for the change in the speed of RNA replication was**

   a. a change in the replicase.
   b. a change in reaction conditions among tubes.
   c. a change in the RNA.

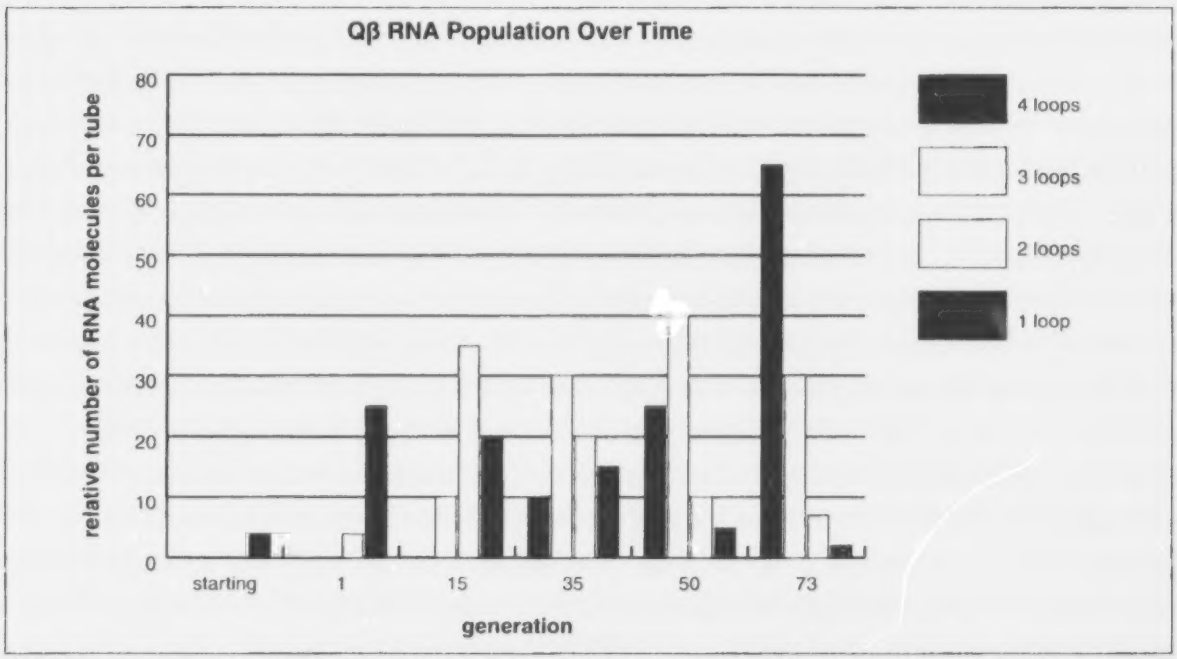


Figure 3.7 Summary graph of simulated results from the *in vitro* evolution experiment.

c. A change in the RNA. All reaction conditions were the same for each tube throughout the experiment. Each tube had the same amount of identical replicase enzyme.

3. **Two major changes occurred in the RNA molecules during the course of the experiment that account for their altered replication speed. First, copying errors that replaced one base with another were randomly made by the "sloppy" replicase. The second and major reason for altered replication speed is apparent from the general trend in the size of the molecules across time.**

   a. How would you describe this trend?
   b. What happened to the longer molecules?

   The RNAs of successive generations became progressively shorter. In fact, the molecules that "survived" into the 73rd generation had lost 83 percent of their original length (from 4,000 bases to 680). These shorter molecules initially appeared because the replicase errs further by occasionally making broken, partial copies of the RNA. If a molecule became shorter while retaining its origin of replication sequence, its replication could be completed faster, and thus it had a competitive advantage in an environment that selected on the basis of replication speed. The longer molecules were effectively becoming "extinct" due to their competitive disadvantage.

4. **Imposing a time limit on the population for replication effectively made the copying process a competition, a race. This competition was certainly not "intentional" on the part of the molecules, but it was**

inevitable simply because of the way the experimental "environment" operated. The reason each "generation" of RNA replicated faster than the preceding one is that

a. the faster molecules in each generation produced more "offspring" copies like themselves, thus increasing their chances of being randomly transferred to the next generation.
b. each generation became progressively better "adapted" to the demands of the environment.
c. both are correct.

c. Both are correct. Evolution has no "intended" outcome, but the inevitable result of evolution's three key factors (replication, variation, and selection) acting in concert across time is a progressive change toward better adaptation in members of descendant populations. These better-adapted individuals have a greater effect on the genetic make-up of the next generation.

5. **If you were to test the RNAs of successive test tube generations for their ability to carry out natural infection-replication cycles in bacteria, you would expect to find**

a. increased infectivity.
b. decreased infectivity.
c. no change in infectivity.

b. Decreased infectivity. Infectivity in the test tube system was markedly impaired because there was no dependence on *E. coli* for replication; the RNAs could lose genes necessary for infectivity without being at a disadvantage. Indeed, by shedding bases the molecules were at a competitive advantage.

6. **After going from a length of 4,000 bases to approximately 700 (shedding more than 80 percent of the genome) the molecules became no shorter. The reason is that**

a. RNAs shorter than this were not sampled and transferred to the next tube.
b. the replicase could not copy Qβ RNA molecules shorter than this.

b. The replicase could not copy Qβ RNA molecules shorter than this. Before copying can begin, the replicase must bind to the origin of replication sequence in the RNA. Below a certain size, an RNA molecule is too short to bind the replicase. Seven hundred bases may seem like a large number for the replicase to interact with. In fact, only some of these bases are directly contacted by the replicase, but they are all needed to ensure that the RNA folds into the correct three-dimensional shape to be recognized.

7. **If you were to test the shortened or mutated RNAs in the test tube for their ability to carry out a natural infection-replication cycle in bacteria, you would expect to find**

a. increased infectivity.
b. decreased infectivity.
c. no change in infectivity.

b. Decreased infectivity for *E. coli* because in the test tube system *E. coli* becomes irrelevant to replication and the "survival" of the phage RNA. The experimenter provided free in the tube everything that would be available only inside a bacterial cell in a natural infection cycle. As a result, the chance loss of those sections of the RNA genome involved only with allowing the virus to enter a bacterial cell had no negative consequence on survival in the test tube environment. All that was essential for survival was the correctly folded origin of replication sequence needed to recognize the replicase.

8. **A change in environmental conditions will select for new "traits" in molecules, just as it selects for new phenotypic traits in populations of organisms evolving in nature. Spiegelman demonstrated this fact by changing the test tube environment in various ways and repeating the experiment. In one case, he added to each tube a chemical inhibitor of the replication process. He added just enough inhibitor to significantly slow but not completely prevent replication. Over time, the replication speed of the RNAs**

a. decreased.
b. increased.
c. remained unchanged.

b. Increased. After an initial slowdown in the first few generations, the replication speed of the population progressively increased. The final population of molecules had "adapted" to the new environment and could replicate in the presence of the inhibitor almost as fast as the wild type RNA could in its absence. Random mutations in the RNAs conferred resistance to the inhibitor on some molecules, and these went on to populate later resistant generations. Some resistant molecules grew dependent on the inhibitor and were unable to replicate in its absence. (You might propose a molecular explanation for this observation). Think about the results of these inhibitor studies in light of the growing problem of antibiotic resistance by disease-causing bacteria.

The results of the inhibitor experiment make the important point that "fitness" varies with the environment: Molecules or individuals that are genetically capable under one set of circumstances will not necessarily be capable if those circumstances change. Another general point is that, short of extreme environmental changes, evolution—through the combined effects of replication, mutation, and selection—will produce over time an altered population that is able to cope with the new environment.

Since these classic experiments were conducted, powerful new approaches to studying RNA and DNA in the laboratory have been developed. Techniques such as the polymerase chain reaction (PCR) for *in vitro* replication of nucleic acids, rapid nucleotide sequencing methods, and assays of nucleic acid function provide very efficient ways to study these molecules. Scientists are now asking ever more probing questions about nucleic acid functions, those that already exist in nature and those that might be fashioned in the lab. Today,

for example, it is possible to synthesize in the test tube large populations of RNA or DNA molecules with randomly varying base sequences. Scientists can select from these diverse populations only those molecules that possess a particular desired function, such as the ability to catalyze a chemical reaction or bind specifically to some other molecule. Nucleic acid sequences discovered in this way might one day be used to disable disease-causing viruses and bacteria or even treat certain forms of cancer.

Directed molecular evolution, as this laboratory approach to molecular evolution is called, is simply a more powerful and efficient variation on Spiegelman's early studies of RNA evolution. The same basic evolutionary features of population variation, selection, and replication are employed in both cases. By exploiting the power of such experimental approaches, we can learn much about the range of potential RNA functions and can use this information to develop evidence-based hypotheses for how the first evolving molecules might have come to be. In Exercise 3.4, you will explore a sophisticated approach to molecular evolution.

Additional Information

Mills, D.R., Peterson, R.L., & Spiegelman, S. (1967, July 15). An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proceedings of the National Academy of Sciences, 58*(1): 217–224.

Joyce, G.F. (1992, December). Directed molecular evolution. *Scientific American*, 267(6): 90–97.

Szostak, J.W., & Ellington, A.D. (1993). *In vitro* selection of functional RNA sequences. In Gesteland, R.F., & Atkins, J.F. (Eds.), *The RNA world* (pp. 511–533). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

FOR YOUR INFORMATION
Molecular Selection in Real Time

One of the great difficulties of studying natural selection in the wild is that selection generally takes place across a long period of time. Often scientists must be satisfied with reconstructing earlier events by looking at the results of those events. For example, in the 19th century, Charles Darwin observed a wide range of beak sizes and shapes among various species of songbirds isolated in the Galápagos Islands. The birds all turned out to be finches, yet their diets and habitats were very different, and their beak characteristics correlated to their different ways to get food. Darwin observed these results of natural selective pressure, but he did not actually watch the process happen.

Fortunately, scientists have had opportunities to observe natural selection taking place in the wild and in real time. Darwin's finches provided such an opportunity. Several scientists extended observations of these finches in the 1970s and later. These researchers collected data during several seasons, measuring and recording the average size of beaks in populations of different finch species and the size and availability of the seeds on which these birds fed. The data showed that during the observation period, changes in average beak size occurred in correlation with changes in the food supply of seeds. Dramatically, changes in the characteristics of the finch population could be observed in just one or two seasons.

FOR YOUR INFORMATION (continued)

Recently, scientists carried out experiments with selection in the laboratory, using populations of molecules. These modern *in vitro* selection (or directed molecular evolution) experiments trace their origins to the RNA replication work of Spiegelman. But they are much more powerful, able to generate RNA or DNA molecules with *predictable* shapes and functions. These experiments are based on the following goal: to direct the evolution of a random population of RNA molecules such that the population becomes enriched for molecules with a desired function.

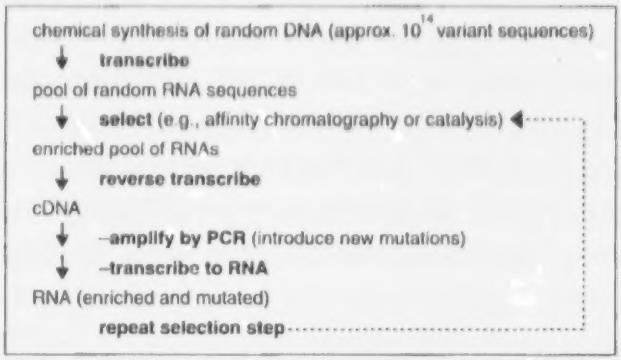

chemical synthesis of random DNA (approx. 10^{14} variant sequences)
transcribe
pool of random RNA sequences
select (e.g., affinity chromatography or catalysis)
enriched pool of RNAs
reverse transcribe
cDNA
–amplify by PCR (introduce new mutations)
–transcribe to RNA
RNA (enriched and mutated)
repeat selection step

Figure 3.8 The strategy of directed molecular evolution. A large, randomly synthesized population of nucleic acids undergoes repeated cycles of selection followed by amplification. The result is a particular sequence well suited to the selection criterion.

For directed molecular evolution experiments (Figure 3.8), scientists start with a large population of random RNA molecules (approximately 10^{15}), each with its own particular base sequence and shape. This starting mixture provides the variation that is required for evolution to take place. The scientists subject this diverse population of molecules to a selective test (also called *selective pressure*) by requiring that molecules possess some specific function, such as binding to a test substance or catalyzing a chemical reaction, before the RNA molecule can be replicated. Molecules able to meet the selective pressure have a reproductive advantage over the others. Selective reproduction (replication) of RNAs passing the test is typically accomplished by the laboratory procedure known as the *polymerase chain reaction* (PCR), in which many copies of the RNA molecules can be generated in the test tube. By repeating this cycle of selection and reproduction of successful variants several times, the final population of RNA molecules is no longer random. Instead it is enriched for particular structures that can carry out the desired molecular function.

Directed molecular evolution is simply a more powerful and efficient variation on Spiegelman's pioneering *in vitro* studies of RNA evolution. Both of these laboratory experiments are nevertheless true examples of evolution at the molecular level. The same basic elements are at play here as in the evolution of a population of living organisms: *variation in the population, selection of individuals based on some essential function or ability,* and *selective reproduction of these individuals.*

In vitro molecular evolution is, of course, much faster than the evolution of complex organisms. But both processes are rather inefficient: Many molecules or organisms must be tested for each one able to successfully meet the test. Nevertheless, the evolutionary process for both cases is powerful across time, being able to generate RNA molecules with new functions and populations of organisms adapted to changing environments.

Ideally, an *in vitro* process would be able to predict the exact RNA base sequence and shape required for a given task and synthesize only this molecule, rather than having to screen many billions of random molecules. Unfortunately, we are a long way from knowing enough about how molecular composition determines shape and function to be able to use this more efficient approach. Nevertheless, as we learn more from the *in vitro* selection experiments about the capabilities of RNA and DNA, we come closer to this goal.

Exercise 3.4: A Nifty Trick with RNA in the Laboratory

A prediction of the RNA world hypothesis is that a self-replicating RNA once existed in nature (and may still exist) and initiated the process of biological evolution. In Exercise 3.2, you saw how, in principle, RNA might employ its catalytic ability to replicate itself, and you examined laboratory approaches, like directed molecular evolution, that can evolve RNAs with specific abilities. The search for a self-replicating RNA continues in the lab and in the field. In this exercise, you will focus on a recent laboratory product of this search, an RNA molecule that neatly unites the ideas of directed molecular evolution, RNA catalysis, and RNA replication.

Procedure

Figure 3.9 depicts an experimental reaction system for the replication of RNA. This system was developed recently in the lab of Gerald Joyce at The Scripps Research Institute. The RNA molecule at the heart of this system functions as both a template molecule and a catalyst. It was originally generated in the laboratory using the technique of directed molecular evolution by screening a large population of 10^{15} randomly synthesized RNA molecules. The researchers were

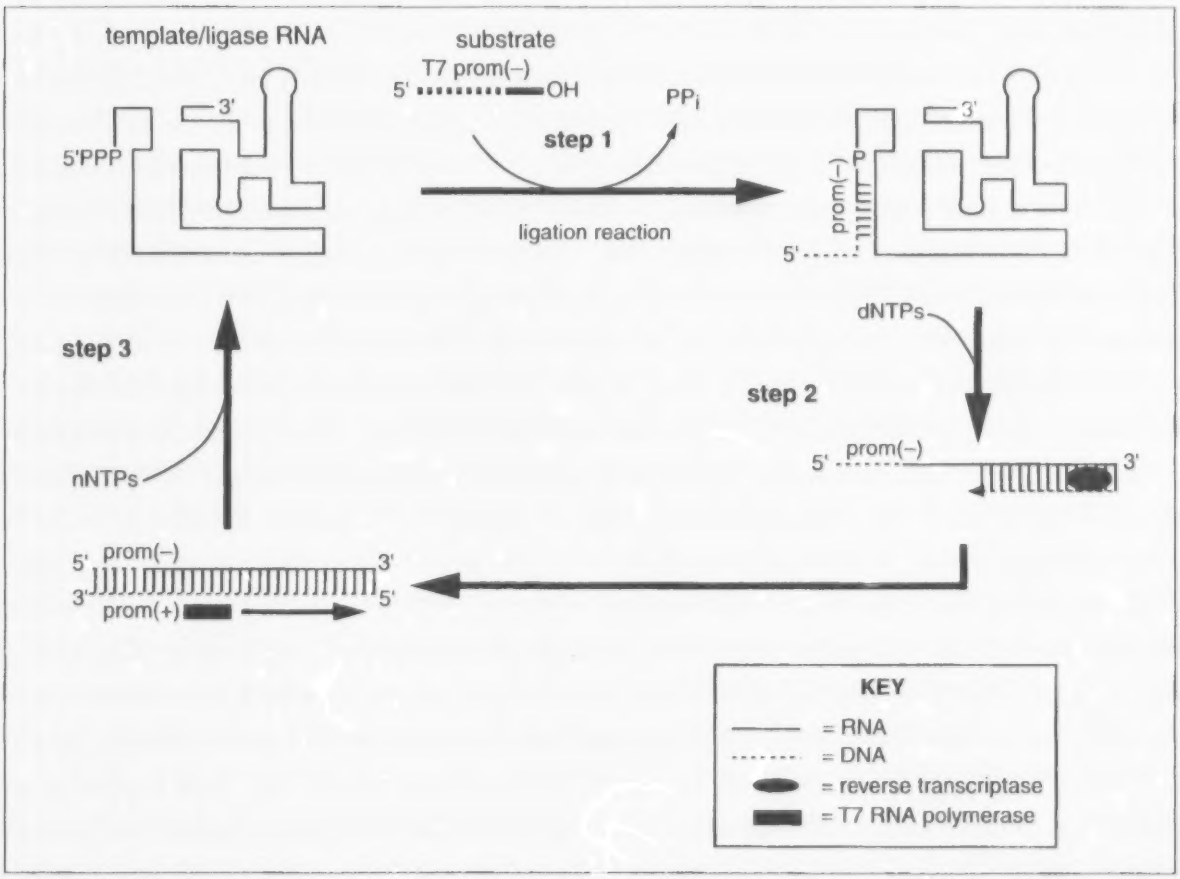


Figure 3.9 *In vitro* system for the continuous replication of RNA. RNA strands are shown as solid lines, DNA strands as dashed lines.

able to select from this varied population an RNA sequence, designated E100, capable of acting efficiently as both template and ligase catalyst in this system.

To understand what is happening in Figure 3.9, apply the templating strategy for replication that you studied in Exercise 3.1 (Figure 3.4). Recall that just as in sculpture, the strategy to reproduce an original object is to make the complement of the original, and then make the complement of that complement: *The complement of the original's complement is the original.* As you work through this reaction sequence and make sense of the individual steps, you will appreciate its elegant internal logic as a simple application of the templating principle. In a later exercise, you will have the chance to perform this replication reaction in the laboratory.

Begin this exercise by analyzing what is happening in Figure 3.9, then answer the Challenge Questions.

Step 1

- The RNA molecule in the upper left of Figure 3.9 is the ribozyme that serves both a template function and a ligase catalyst function in this system. We will refer to it as the *template/ligase*.
- The molecule labeled *T7 prom(-)* is a short single strand made up of DNA nucleotides, with a few RNA nucleotides at its 3′-end. T7 prom(-) is the substrate for the template/ligase, and it encodes one strand of the T7 promotor. Later in the reaction, this promotor will serve, in double-stranded form, as the start site for a transcription event by the protein enzyme T7 RNA polymerase. (Recall that RNA polymerases, in general, transcribe DNA templates to make RNA molecules.)
- In Step 1, the template/ligase uses its template ability to properly align itself with complementary bases in the substrate; it then uses its catalytic ability to ligate the substrate to its own 5′-end.

Step 2

- The DNA primer is a short sequence of DNA that is complementary to the 3′-end of the template/ligase RNA (shown hybridized in Step 2; the RNA is shown unfolded at this point for simplicity). The primer's sole function in the reaction is to provide a start site for the protein enzyme reverse transcriptase. Recall that reverse transcriptase copies an RNA template (or a DNA template) into a DNA strand.
- The result of the reverse transcriptase step in this reaction is a double-stranded molecule that is a hybrid. One strand of the hybrid is the original ribozyme RNA with its attached prom(-) portion of the T7 promotor; the newly made second strand is DNA that is complementary to the ribozyme (and also provides the second strand of the T7 promotor).

Step 3

- T7 RNA polymerase uses the now-functional (double-stranded) T7 promotor as its start site to carry out a transcription reaction.

Challenge Questions

1. **What is accomplished in Step 1 of the reaction sequence? Which function(s) of the RNA ribozyme are employed at this step?**

   The ribozyme ligates the substrate molecule (encoding one strand of the T7 promotor) to its own 5′-end. The ribozyme acts as both a template (by base pairing with the substrate to align it properly) and a catalyst (by ligating the DNA tail to its 5′-end.)

2. **A double-stranded molecule results from Step 2, reverse transcription. What is the coding relationship of the newly synthesized second strand to the RNA ribozyme sequence? Which part of the templating strategy is accomplished in Step 2? At the completion of this step, what function can the T7 promotor in the double-stranded molecule perform?**

   Reverse transcriptase synthesizes a DNA strand that is complementary to the ribozyme RNA sequence. In terms of the templating strategy, Step 2 is the creation of the original's complement. The double-stranded T7 promotor in the hybrid molecule is now a functional start site for T7 RNA polymerase in the next step.

3. **In terms of the templating strategy, what does RNA polymerase accomplish in Step 3? What is the product of this step? What effect does this step have on the number of template/ligase RNA molecules in the tube?**

   RNA polymerase produces the complement of the original's complement and thus makes a new copy of the original template/ligase ribozyme. The template/ligase RNA is being replicated and increasing in concentration.

4. **Assume there is a surplus of monomer building blocks (dNTPs and rNTPs) and substrate molecules in the tube. What will happen across time? If an aliquot of this reaction mixture is transferred to a new tube containing monomers and substrate molecules, what would you expect to happen?**

   Because the reaction sequence is a continuous cycle, it will repeat over and over in the tube, generating ever more RNA. Transferring an aliquot to a fresh tube of monomers and substrate allows the reaction to keep running. In principle, this serial transfer process could be continued indefinitely.

5. **The enzymes employed in the system, especially RNA polymerase, have relatively high error rates (approximately one to two errors per molecule copied). Comment on this system's potential to generate novel RNA sequences in the laboratory.**

   The system continuously generates variants of the original sequence as the RNA is replicated and can yield new sequences with novel properties.

6. **The Joyce system, like the Qβ RNA replication system developed by Spiegelman, is able to reproduce RNAs continuously. Both of these systems, however, employ reaction components that prevent them from qualifying as the true self-replication of RNA. What are they? Which system is closer to true self-replication? Why?**

> Both systems rely on protein enzymes to generate copies of the RNA. In the Qβ system, replicase was needed to copy RNA to RNA; in the Joyce system, reverse transcriptase (RNA to cDNA) and RNA polymerase (cDNA to RNA) were required. True self-replication requires that the RNA carry out the steps of replication on its own, without the aid of proteins. The Joyce system is closer to self-replication in the sense that it relies on the RNA's catalytic ability to get the ball rolling, a feature not present in the Qβ system. Researchers are using directed evolution methods such as this to find such a truly self-replicating RNA.

Model *in vitro* selection by developing a variation of the "ball test" and physical models of RNA from Exercise 1.3 to have students simulate an *in vitro* selection experiment at their desks.

The experimental system for RNA replication featured in Exercise 3.4 (see Figure 3.9) nicely brings together the ideas of directed molecular evolution, RNA catalysis, and RNA replication. The system was developed recently in the laboratory of Gerald Joyce at The Scripps Research Institute (Wright & Joyce, 1997). This exercise encourages students to understand and analyze this seemingly complex reaction system as an example of the simple templating strategy studied in Exercise 3.2 (see Figure 3.4): *The complement of the original's complement is the original.*

The RNA molecule at the heart of this system was originally generated in the laboratory using the technique of directed molecular evolution. The system's ability to replicate RNA takes advantage of two important abilities of the RNA: to act both as a template and as a catalyst. As a template, the RNA first pairs and aligns itself with a short, complementary, single-stranded substrate of DNA that encodes the promotor site for T7 RNA polymerase, a protein enzyme also present in the tube that will eventually make a copy of the RNA. As a catalyst, the RNA next ligates the 3′-end of the substrate to its own 5′-end. (The substrate bears four RNA ribonucleotides at its 3′-end to allow this ligation.) Following an intermediate reverse transcription step, in which a DNA complement of the original RNA is made, T7 RNA polymerase transcribes this complement to replicate the original catalytic RNA.

This replication system is continuously self-sustaining because the tube contains all of the materials necessary for the ligation-replication cycle to repeat over and over (such as NTPs and dNTPs, substrate, primer DNA, and the protein enzymes RNA polymerase and reverse transcriptase). The result is many copies of the original RNA molecule. In principle, this reaction can be made to continue indefinitely, simply by periodically transferring an aliquot of the initial reaction to a new tube containing fresh monomer building blocks and the other essential factors. The RNA will inevitably evolve across time in such a system.

After the students have worked through the exercise and answered the questions on their own, you may want to discuss and explore the system further. You might explain that amplification of the original RNA occurs at two points in the cycle. First, the DNA-RNA hybrid intermediate can be transcribed over and over in the tube to generate many new RNA copies. Second, each RNA copy can then carry out another ligation reaction and initiate another round of replication in the tube. Also, be sure that students see the utility of such a continuous replication system for studying the evolution of RNA in the laboratory.

Note that once the nucleotide monomers and enzymes are in the reaction tube, replication could be initiated by adding either the purified catalytic RNA (not practical due to the long-term instability of purified RNA) or by adding the double-stranded intermediate. In the laboratory protocol in Exercise 3.5, students initiate the reaction by adding the more stable double-stranded molecule (referred to in the protocol as *input PCR DNA*, because it must be generated before the experiment by PCR amplification from the recombinant plasmid [E100-3], which encodes it; see Wright & Joyce, 1997). The necessary PCR primer sequences are also shown (see Oligonucleotide Reagents).

Exercise 3.5: Continuous *in vitro* Replication of RNA

Here is your opportunity to carry out in the lab the RNA replication reactions that you explored in Exercise 3.4. Your teacher will provide you with the following reagents:

The following Reagent List contains teacher preparation instructions. The *Student Pages* have a modified list.

REAGENT LIST

a. **Tube 1, stock mix*** - Prepare this tube prior to class; each stock tube is enough for one time course experiment.

218.2 µl high-quality distilled water	
20 µl KCl (1 M)	final concentration during reaction = 50 mM
12 µl EPPS buffer (1 M, pH 8.5 at 22°C)	30 mM
10 µl $MgCl_2$ (1 M)	25 mM
8 µl Spermidine (100 mM)	2 mM
20 µl dithiothreitol (DTT) (100 mM)	5 mM
8.4 µl cDNA primer, TAS 1.23** (100 mM)	2 µM
32 µl ribonucleoside triphosphate (rNTP) mix*** (25 mM in each rNTP)	4 x 2 mM
3.2 µl deoxy-NTP (dNTP) mix**** (25 mM of each dNTP)	4 x 200 µM

* All reagents must be ultrapure grade; this is important to success.

** Have this and the two other required oligos synthesized commercially according to the sequences provided in Oligonucleotide Reagents, following.

*** Prepare rNTP mix ahead by adding an equal volume of each of the four individual rNTPs (each at 100 mM) to a tube. A set of four filter-sterilized aqueous solutions, each containing 100 mM of rCTP, rUTP, rATP, or rGTP, is available from Boehringer Mannheim (catalog # 1277057).

**** Prepare dNTP mix as above, using the four dNTPs (each at 100 mM). Also available from Boehringer Mannheim (catalog # 1277049).

b. **Tube 2, 22 µl substrate (S162-2) (2[illegible] µM)** - Have this 35-mer mixed oligo synthesized commercially according to the sequence provided in Oligonucleotide Reagents. The oligos must be purified on a 20% acrylamide gel.

c. **Tube 3, 35 µl T7 RNA polymerase (100 U/µl)** - Use ultrapure grade; aliquot as supplied. Promega, catalog # P4074, 80 U/µl 10,000 units.

d. **Tube 4, 35 µl M-MLV reverse transcriptase (200 U/µl)** - Use ultrapure grade; use as supplied. Promega, catalog #M1705, 50,000 units, or Amersham-Pharmacia, catalog # E70456Z, 100,000 units.

e. **Tube 5, 2.6 ml oxazole yellow (YO-PRO-1) dye (1 mM)** - Available as 1 mM stock in DMSO from Molecular Probes, Inc. (catalog # Y-3603). Use as supplied. Protect from light.

f. **Tube 6, 54 ml input PCR DNA (16 nM)** - Have the input PCR DNA prepared commercially by PCR from *plasmid E100-3*. This recombinant plasmid contains a copy of the T7 RNA polymerase promotor upstream of a DNA insert whose template strand is the complement of the catalytic RNA used in this experiment. The template strand of the input PCR DNA is copied by RNA polymerase in this system to generate new copies of the catalytic RNA. For copies of the plasmid E100, see Wright & Joyce (1997).

OLIGONUCLEOTIDE REAGENTS

The three oligos required for this experiment can be custom ordered from companies that specialize in synthesizing custom oligos (for example, Operon or NBI). TAS 1.23 and TAS 2.54 are required as primers for the PCR amplification of the plasmid insert; TAS 1.23, but not TAS 2.54, also functions as a reagent in the replication reaction mix. The third required oligo, S162-2, is a mixed deoxyribo-ribonucleotide and serves as the substrate for the ribozyme in the replication reaction mix. *All of the oligos should be purified on a 20% acrylamide gel prior to use.* The oligo sequences are shown below:

TAS 1.23: 5′-d (GCTGAGCCTGCGATTGG)-3′ (approximately 1 µmole)
TAS 2.54: 5′-d (CTTGACGTCAGCCTGGA)-3′ (approximately 1 µmole)
S162-2: 5′-d (CTTGACGTCAGCCTGGACTAATACGACTCAC) **r (UAUA)**-3′ (approximately 1 µmole)

Procedure

You will use a dye that fluoresces when bound to the nucleic acid to visualize the production of RNA across time.

CAUTION: Always wear eye protection when viewing.

1. **Place tube 1, tube 6, and a spectrofluorimeter cuvette (0.5 ml capacity) at 37°C.**

2. **To tube 1 at 37°C, add the following, in order:**

   16 µl of T7 RNA polymerase from tube 3
   16 µl of M-MLV reverse transcriptase from tube 4
   10 ul substrate (S162-2) from tube 2
   1.2 µl of oxazole yellow dye from tube 5

3. **To tube 1, add 25 µl (400 fmol) of prewarmed input PCR DNA (from tube 6).**

4. **Immediately transfer this mix to the prewarmed cuvette and begin taking spectrofluorimeter readings (excitation wavelength = 491 nm; emission wavelength = 509 nm).**

5. **Run the reaction in the cuvette at 37°C for 1 hour, taking fluorescence readings every 2 minutes. As you monitor the reaction, remind yourself of what is occurring in the tube by reviewing Figure 3.9 in Exercise 3.4.**

6. **Plot fluorescence readings versus time to observe the RNA growth curve.**

Challenge Questions

1. **How did the level of fluorescence emitted by the reaction change across time? How do you account for this change?**

2. **How might this system of continuous replication be applied in a directed molecular evolution type of experiment?**

This qualitative version of the continuous replication experiment does not require a spectrofluorimeter. Instead, a UV light box or hand-held UV light is used for viewing. As the RNA replication proceeds, students remove aliquots of the reaction at different times and quench the reaction in the aliquots by adding each to a separate tube containing the fluorescent dye. Students view the assembled time series of tubes by UV (or ideally 490 nm light) to observe the progressive increase in fluorescence over time. The changes in fluorescence intensity when viewed by eye are detectable but subtle (see Figure T3.1). Consequently, the impact of the result is somewhat less than in the quantitative version of the experiment.

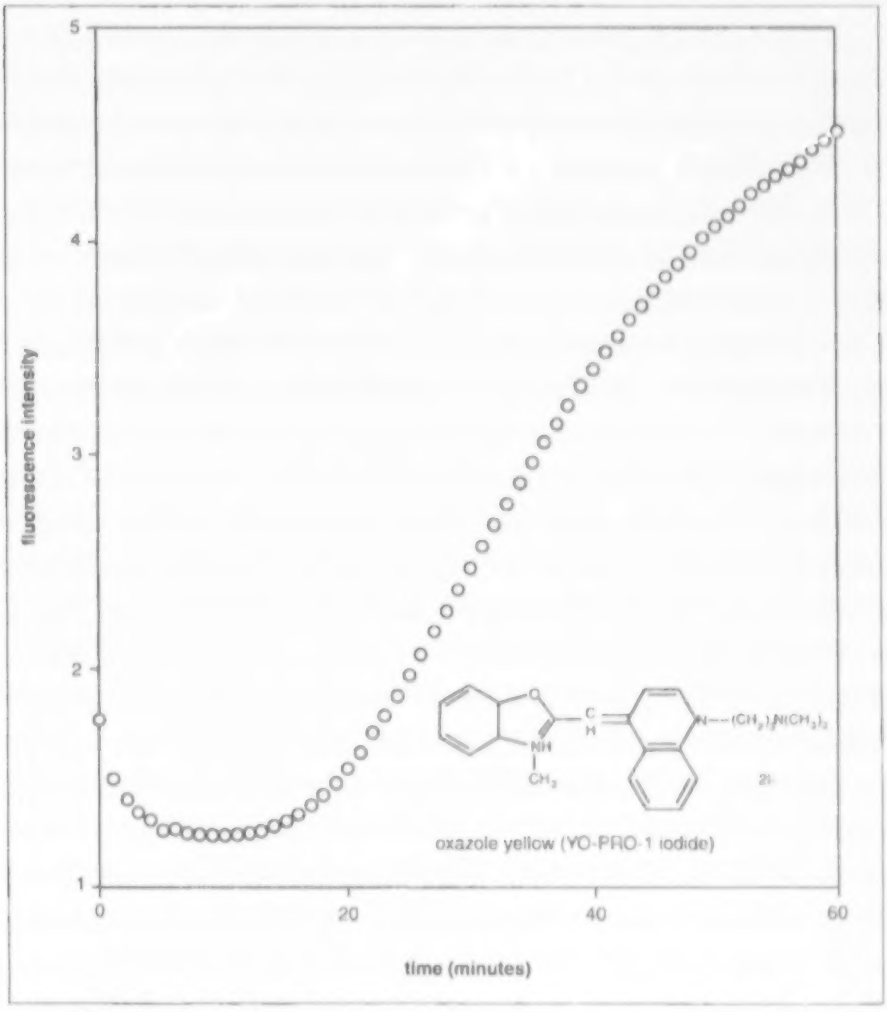


Figure T3.1 Fluorescence emission curve during continuous *in vitro* replication of RNA.

If time and resources permit, conduct this extension after the initial reaction at the 60-minute interval. Have the students remove an aliquot of the reaction from the cuvette and dilute it 1:500 with distilled water. Transfer 25 µl of this diluted sample to a fresh tube of stock mix with the added T7 RNA polymerase, reverse transcriptase, substrate, and dye, as in Step 2 above. Transfer this new reaction mix to the warm cuvette and begin monitoring fluorescence once again. This will allow the students to observe the reaction take off again, populating the fresh tube with more newly synthesized RNA. Note: The input PCR DNA need not be added to this new tube and any subsequent tubes because it is provided by the diluted aliquot the students transferred.

You might ask your students what component of the aliquot transferred to the new tube is capable of seeding the production of more RNA. In fact, two capable components are present: copies of the ribozyme RNA and the double-stranded DNA-RNA hybrid that gives rise to these RNAs.

You might have the students predict the effect of omitting one of the reagents from the mix (for example, substrate S162-2, T7 polymerase, or the rNTPs) and then have them test their prediction experimentally.

Alternative Exercise 3.5: Continuous *in vitro* Replication of RNA Using Visual Detection of Fluorescence

Here is your opportunity to carry out the RNA replication reactions that you explored in Exercise 3.4. Your teacher will provide you with the following reagents:

The following Reagent List contains teacher preparation instructions. The *Student Pages* have a modified list.

REAGENT LIST

a. **Stock mix** - Prepare this stock mix ahead of time. Aliquot and evaporate as described below.

16 µl KCl (1 M)	final concentration during reaction = 50 mM
9.6 µl EPPS buffer (1 M, pH 8.5 at 22oC)	30 mM
8 µl MgCl2 (1 M)	25 mM
6.5 µl Spermidine (100 mM)	2 mM
16 µl dithiothreitol (DTT) (100 mM)	5 mM
26 µl nucleoside triphosphate (NTP) mix (25 mM of each NTP)	4 x 2 mM
2.6 µl deoxy-NTP (dNTP) mix (25 mM of each dNTP)	4 x 200 µM
6.5 µl cDNA primer - TAS 1.23 (100 µM)	2 µM
229.2 µl high purity distilled water	
320.4 µl total volume	

Measure 20 µl aliquots of the stock mix into 16 1.5-ml Eppendorff tubes.
Evaporate to dryness in a speed-vac or comparable evaporator.
Store at -20°C until the lab period. This provides enough reagent for 8 groups of 3 to 4 students to replicate the experiment twice.

BEST COPY AVAILABLE

b. **Tube 1, dried 20 µl aliquot of stock mix** - Provide each group of students with *two* each of tube 1. Also give each group *one* each of tubes 2 to 6; each of these tubes has 2X volume of reagent, enough to replicate the experiment twice.

c. **Tube 2, 5 µl substrate (S162-2) (25 µM)** - The working concentration of this reagent (25 µM) is one-fourth of that used in the quantitative experiment. This and the other two required oligos (primers TAS 1.23 and TAS 2.54) may be synthesized commercially, as in the quantitative experiment.

d. **Tube 3, 4 µl T7 RNA polymerase (80 U/µl)** - Use ultrapure grade; aliquot as supplied. Promega, catalog # P4074, 80 U/µ.

e. **Tube 4, 4 µl M-MLV reverse transcriptase (200 U/µl)** - Use ultrapure grade; aliquot as supplied.

f. **Tube 5, 160 µl oxazole yellow dye (YO-PRO-1) (5 µM)** - The working concentration of dye (5 µM) is 1/200th that of the quantitative experiment. Dilute in distilled water. Protect from light.

g. **Tube 6, 24 µl input PCR DNA (16 nM)** - Have this DNA prepared commercially by PCR from *plasmid E100-3* (see Wright & Joyce, 1997).

Procedure

You will use a dye that fluoresces when bound to the nucleic acid and illuminated with UV light to visualize the increase in RNA across time.

CAUTION: Always wear eye protection when viewing.

1. **Label six half-milliliter Eppendorff tubes as follows: 0, 5, 10, 15, 20, and 25 minutes.**

2. **Put the tubes on ice; to each, add 13 µl of oxazole yellow dye from tube 5.**

3. **Place tubes 1 (dried stock mix) and 6 (input PCR DNA) at 37°C.**

4. **Add the following to tube 1 at 37°C:**

   - **5.4 µl distilled water**
   - **2.0 µl of substrate (S162-2) from tube 2**
   - **1.3 µl of T7 RNA polymerase from tube 3**
   - **1.3 µl M-MTV reverse transcriptase from tube 4**

5. **At time zero, add 10 µl of prewarmed input PCR DNA to tube 1. Mix and immediately remove a 2 µl aliquot from this tube and transfer it to the iced dye tube labeled "0."**

6. **At 5-minute intervals, transfer a 2 aliquot from tube 1 to the appropriately labeled dye tube.**

7. **Visualize the six dye-containing tubes as a group by UV light or, ideally, 490 nm light.**

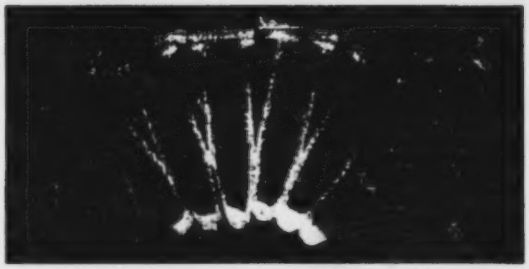


Figure T3.2 A series of aliquots from the continuous replication reaction viewed by UV shows increased fluorescence across time.

Challenge Questions

1. **How did the level of fluorescence emitted by the reaction change across time? How do you account for this change?**

2. **How could this system of continuous replication be applied in a directed molecular evolution type of experiment?**

FOR YOUR INFORMATION
What Airplane Design Hasn't Got to Do with Biology

When we look at the effects of natural selection on a population of organisms or of molecules, we see that certain traits become more widespread in the population if they provide some advantage under a particular set of circumstances. This phenomenon is commonly known as survival of the fittest, and it is easily observable. It is tempting, then, to think that whatever characteristic or individual comes to dominate a population must be the best structure possible to carry out the needed functions. This view is in error. When we look at a living organism or even at individual molecular structures, we realize that an engineer would have designed them differently.

Many people are under the impression that machines evolve just like living organisms, but this is hardly the case. Consider a simple example, the human-designed airplane. The Wright Brothers achieved the first successful powered air flight by building a biplane. It had a wooden frame, a cloth skin, an open cockpit, and a propeller driven by a reciprocating piston engine. This frail craft would not be much competition for a large modern passenger plane carrying 300 times as many passengers and flying 100 times faster at much higher altitudes.

Although we may say casually that one plane "evolved" into the other, in fact, what really happened over the years was that many new prototype planes were created. But each new plane was designed, tested, and built on the ground before it was flown. Understandably, engineers choose not to take the risks associated with construction on-the-fly. Living systems do not have this luxury.

A living system has to keep going, keep reproducing, without missing a generation, or it disappears. Cells do not get to take time out while the machinery for protein synthesis or replication is replaced by a completely new and better design. This would be analogous to the old parts of an airplane being dismantled in midair while new parts of better design were constructed and attached without ever landing or crashing. Intermediate structures, such as a plane with one wooden wing and one new metal one, would have to be sufficiently airworthy for uninterrupted flight.

FOR YOUR INFORMATION (continued)

Because life comes from pre-existing life, living organisms must in effect evolve in "midair." The fictional Dr. Frankenstein went into the laboratory and constructed a new human from spare parts before he sparked it into life, but in the real world of living systems, life is continuous. New organisms must be built from the blueprints of their parents without interruption.

That biological evolution must take place on-the-fly is an enormous constraint. The feasible alterations to an airplane in flight are much more limited than the experiments and changes that engineers can carry out on the ground, in the safety of the hangar. The engineer builds painstakingly with planning and foresight; evolution meanders along testing randomly generated variations. The combined effects of trial and error plus natural selection take an unavoidable toll in failed experiments, but inevitably lead to better designs for the survivors. Importantly, what survives in nature will not necessarily be the best design possible but simply the most advantageous structure available at the time from a large pool of variants.

Activity 4
RNA Evolution in Health and Disease

Activity at a Glance

Exercise 4.1:	Students learn how rapid growth rate and large population size play a role in the generation of antibiotic-resistance mutations in bacteria.
Exercise 4.2:	Students learn how the inappropriate use of antibiotics promotes the evolution and emergence of resistant bacterial strains.
Exercise 4.3:	Students conduct a laboratory experiment in which they shape the evolution of a bacterial population by selecting for a streptomycin-resistant mutant.
Exercise 4.4:	Students consider the evolutionary basis for drug resistance in RNA viruses by analyzing the evolution of HIV in the microcosm of a single patient.
Exercise 4.5:	Students consider the molecular basis for the evolution of viral drug resistance and compare the RNA of different viral strains to establish a phylogeny among them.
Extension Exercise:	Students compare RNA sequences to establish phylogenetic relationships among organisms.

Concepts

- Mutations at the DNA and RNA level arise frequently in large, rapidly reproducing populations.
- The emergence of antibiotic resistance among bacteria is an evolutionary process that is promoted by inappropriate use of antibiotics.
- Selection and evolution can be readily observed in the laboratory as well as in nature.
- The evolution of RNA viruses is particularly rapid and poses a significant public health challenge.
- The similarities between RNA and DNA can be used to help establish the ancestor-descendant relationships among all organisms.

BEST COPY AVAILABLE

Focus

Students explore the *in vivo* evolution of RNA.

Connections to Lecture Topics

Use Activity 4 in conjunction with a study of the molecular aspects of disease or as a part of a study of evolution and natural selection.

Estimated Time

2 hours of class time and 2 laboratory periods

Materials Preparation

See annotations and materials list in Exercise 4.3 for laboratory materials.

ANNOTATED STUDENT ACTIVITY

Introduction

A news article carried the following ominous headline:

"Overprescribing: Misuse of Antibiotics Creates Superbugs"
The Salt Lake Tribune
September 17, 1997

The article went on to report:

> **Faced with patients demanding medicine for coughs, congestion and sniffles, doctors wrote 12 million [inappropriate] prescriptions for antibiotics to U.S. adults in a single year—even though the drugs are worthless for colds and other viral infections. Such misuse fuels the spread of bacteria that are resistant to antibiotics, leaving fewer effective drugs for patients with serious bacterial infections, University of Utah physician Merle A. Sande and Colorado doctors reported today in *The Journal of the American Medical Association*. "It's extremely serious," said Sande, the university's chairman of internal medicine. "We are losing all these antibiotics. Our future and our children's future is going to depend on us being more selective."**

No doubt you have encountered similar news stories warning of the overuse of antibiotics and its contribution to the emergence of resistant strains of bacteria. Since 1943, when penicillin was first introduced as the "magic bullet" for curing many infectious diseases, more than 100 additional antibiotics have been developed. Despite this seemingly large arsenal, the unfortunate fact is that most of these agents are becoming less and less effective as widespread misuse promotes the development of resistance among bacterial species. Today, it is estimated that 90 percent of all staphylococcus strains are penicillin-resistant, and several other pathogenic species, such as *Streptococcus*, *M. tuberculosis*, and *P. aeurginosa*, have developed strains resistant to all but a few remaining drugs. Inappropriate use of antibacterial agents is a problem not just in medical practice but in animal husbandry, where low levels of antibiotics are routinely included in livestock feed, and increasingly in consumer products such as soaps and detergents that incorporate antibacterial agents.

In this activity, you will see that the acquisition of antibiotic resistance by bacteria, and of resistance to antiviral drugs by viruses, is a predictable result of the evolution of these organisms' genetic systems.

When you examined RNA replication in Activity 3, you saw that evolution, whether of populations of molecules in the test tube or organisms in the wild, is driven by three basic processes: (1) random generation of mutations in DNA or RNA; (2) replication of these mutations during nucleic acid synthesis, with some of the offspring inheriting these mutations, and (3) enhanced reproductive success (by natural or artificial selection) of those individuals carrying mutations that are advantageous. Now you will explore how these three natural processes can conspire to promote the spread of resistance among pathogens.

Exercise 4.1: Setting the Stage for Antibiotic Resistance

Random mutations are always arising within populations. Most mutations are harmful to their carrier, some are advantageous, and others are "selectively neutral," having no effect on the carrier's reproductive success. Although mutations are inevitable, most organisms manage to keep the *mutation frequency* (the number of mutations per base per generation) relatively low by repairing most of them before they have a chance to be passed on. Thus, the chance of a *particular* mutation occurring in a *particular* individual is low. However, mutations appear often in rapidly growing populations consisting of many individuals.

Consider the bacterial mutations that result in resistance to the aminoglycoside antibiotic streptomycin. Aminoglycoside antibiotics inhibit protein synthesis in prokaryotes (bacteria) by binding to a specific sequence of bases in their 16S ribosomal RNA (see FYI essay *How Aminoglycoside Antibiotics Target Bacterial RNA*). Particular single-base (or point) mutations in the bacterial rRNA can prevent the binding of streptomycin, enabling the bacterium to become resistant to the drug. The likelihood that any particular base will be mutated to cause resistance is small, occurring in perhaps 1 in 1 billion genome replications (that is, the mutation frequency is 1/1,000,000,000 or 10^{-9} per individual). But because bacteria reproduce so frequently (every 20–30 minutes), an infected wound may easily contain billions of individuals. The population in the wound can become large enough that it is virtually certain that at least one bacterium will have acquired streptomycin resistance due to this mutation.

To calculate the probability of finding a particular point mutation in a population of a given size, we would multiply the mutation frequency per base pair per organism (10^{-9}) by the number of organisms present. For example, in a population of 1 billion (10^9) bacteria, the probability equals unity ($10^{-9} \times 10^9 = 1$). That is, the presence of one streptomycin-resistant bacterium is statistically a near certainty by the time the population reaches 1 billion. While one resistant bacterium in a billion might seem like a harmless drop in the bucket, this solitary resistant cell (unlike its millions of susceptible neighbors) would continue to multiply in the presence of streptomycin, producing resistant descendants. Once this mutation has spread in the bacterial population,

treatment with streptomycin and certain other related compounds would prove futile.

Another important factor affecting the rate at which bacteria acquire resistance mutations is the fact that bacteria are quite promiscuous. They have several ways of acquiring and exchanging genetic information, with other members of their own species as well as with other bacterial species. These genetic exchange mechanisms include sexual transfer (conjugation), direct uptake of free-floating DNA (transformation), and transfer from viruses (transduction). Such promiscuous genetic exchange allows for the "accelerated evolution" of resistance in bacteria. (The evolution is accelerated in the sense that a given species can bypass the mutation process by acquiring an already mutated resistance gene from another species.)

Challenge Questions

Read the introduction to Exercise 4.1, then answer the following questions:

1. **Infecting bacteria typically divide once every 30 minutes. If a single bacterium is introduced into a wound at time zero, how many hours would it take for this bacterium to exceed 100 descendants? 1,000 descendants? 10,000 descendants?**

   It would take 3.5 hours to produce 100 descendants, 5 hours for 1,000 descendants, and 7 hours for 10,000 descendants.

2. **Is it possible that a bacterial population size could reach 1 billion *without* a streptomycin-resistant individual arising? Could it reach 3 billion? Explain.**

   Yes, there is a statistical chance. Because mutations are random, a particular mutation never *has* to occur; it only becomes more *likely* to occur as the population expands.

3a. **Is it more likely or less likely that the same mutation would arise independently in two different bacterial cells within a population? Explain.**

   Much less likely. The frequency of this event is $10^{-9} \times 10^{-9} = 10^{-18}$ (that is, 1 in 10^{18} cells or 1 billion times less likely).

3b. **In reality, there are several sites within the bacterial rRNA that can undergo single point mutation and can individually cause resistance to streptomycin. Each of these sites has a roughly equal likelihood of incurring a mutation (approximately 10^{-9}). What effect does the presence of several potential sites for mutation have on the overall chance of streptomycin resistance arising in the bacterial population?**

   The probabilities are additive in this case, and thus increase the total probability that the rRNA will experience a resistance-causing mutation somewhere in its sequence. In contrast, the probability that any two particular mutations will occur *together* in the rRNA is a multiplicative function, and thus is much smaller than either of the individual probabilities.

4. **When antibiotics are required to combat an infection, it makes sense to start therapy as soon as possible to cure the disease quickly and minimize suffering. Explain the evolutionary rationale for starting therapy earlier rather than later.**

   Starting therapy early with an adequate drug dose minimizes the chance that the bacterial population will grow large enough to experience resistance mutations.

FOR YOUR INFORMATION
How Aminoglycoside Antibiotics Target Bacterial RNA

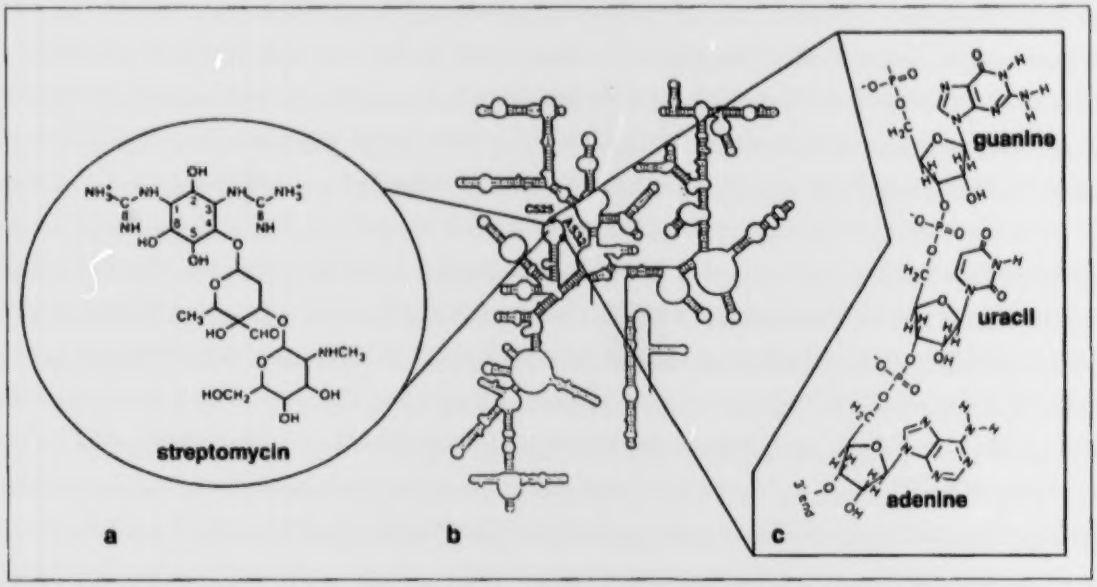


Figure 4.1 (a) The structure of streptomycin. Note the positive charges. (b) The predicted secondary structure of bacterial 16S rRNA, showing the location of the "530 stem-loop" involved in streptomycin binding. The positions of two bases—A532 and C525, whose mutation can confer streptomycin resistance—are indicated. Note that the RNA is shown at smaller scale than the streptomycin molecule. (c) An enlargement of three rRNA bases in the streptomycin binding region, including A532. Note the negative charges in the phosphate backbone.

Review Figure 4.1, then answer the following questions:

5. **What can you hypothesize from Figure 4.1 about the type of chemical interaction that contributes to the binding between streptomycin and the rRNA?**

   The interaction is electrostatic +/-.

6. **Notice in Figure 4.1b that bases A532 and C525 are unpaired and located in the loop region of the binding stem-loop. Speculate how a mutation from A to G at position 532 could alter the binding of streptomycin and promote resistance.**

This new G might pair with a C elsewhere in the same loop. The pairing could distort the loop's shape, decreasing the affinity of streptomycin binding.

7. **The "530 stem-loop" in 16S rRNA is highly conserved among bacteria, varying very little in sequence between bacterial species.**

   a. What does this fact imply about the importance of this region in normal bacterial ribosomal function? Explain.

   This region is quite important in normal bacterial ribosomal function because a little perturbation interferes with optimal function.

   b. Speculate why the ability to produce agents like streptomycin was strongly selected for among certain fungi (genus *Streptomyces*) that compete with both bacteria and other fungal species in the environment.

   Having a powerful antibacterial agent gave these fungi a competitive advantage.

   c. Speculate why streptomycin has no effect on eukaryotic protein synthesis.

   The base sequence and shape of eukaryotic rRNA is different from that of prokaryotic rRNA.

Exercise 4.2: Selecting for the Emergence of Resistance

The ideal effective antibiotic therapy kills all of the infecting organisms before their rapid division can give rise to even a single highly resistant cell. Unfortunately, this ideal is undermined when the invading population is exposed to a dose of the drug that is too low. The trick is for the patient to start therapy as soon as possible and take enough medication across several days to allow tissue levels of the drug to reach a concentration capable of killing all members of the population.

But the ever present genetic diversity in biological populations complicates this task. Recall that for the vast majority of genetic traits, natural populations have a diversity of phenotypes that display the traits to greater or lesser extent. This is also true for the trait of drug resistance in bacteria. Genetic diversity for antibiotic sensitivity within a bacterial population results in individuals with susceptibilities that range from high to low; greater concentrations of the drug are required to kill the less susceptible individuals.

If too little antibiotic is prescribed or if the patient stops taking it before completing the treatment course, tissue levels of the drug never become high enough to kill all the bacteria. Low levels of streptomycin, for example, merely slow the growth of bacteria but do not kill them. Such incomplete treatment is an opportunity for bacteria to evolve into a more drug-resistant population. Under these conditions, the antibiotic is said to exert selection pressure on the population. That is, the antibiotic acts as an environmental factor that

allows a formerly uncommon phenotype (less susceptible) to multiply while inhibiting the previously dominant phenotype (susceptible). Low levels of the drug may select *against* the majority of the population, which is highly susceptible, by inhibiting or killing it. But these low levels end up selecting *for* the minority of the cells, which are less susceptible, by failing to kill it. These less susceptible bacteria continue to divide and make up an ever increasing portion of the evolving population.

Any new random mutation in this modified population that happens to produce an even higher level of resistance will quickly spread in the same way. The inevitable outcome is a population of highly resistant descendants. Further treatment with this drug will now be ineffective, even at higher doses. The only hope at this point is to try another antibiotic, a diminishing option as organisms continue to evolve resistance to ever more agents.

Inappropriate antibiotic treatment promotes the spread of resistance in pathogen populations in another way: It inhibits or kills normal resident bacterial cells that happen to be sensitive to the drug used. This treatment changes the ecology of the wound "environment" and effectively decreases the competition between bacterial species, which would otherwise act to slow the growth of the pathogen. Making matters worse, it is possible that the resistant pathogen can transfer its resistance gene(s) to other bacterial species, both normal flora and other potential pathogens.

Challenge Questions

Read the introductory text to Exercise 4.2, then answer the following questions:

1. **Explain the evolutionary justification for the following statement: The routine addition of antibiotics to livestock feed is ill advised. Can you think of a recent example of a resistant pathogenic bacterial strain that originated in livestock?**

   By routinely exposing animals to low doses of antibiotics, we introduce a selective agent that drives the evolution of their bacterial populations toward resistance. Some strains of the particularly pathogenic *E. coli* 017H7, which is responsible for human deaths from contaminated and undercooked beef, have developed resistance to certain antibiotics.

2. **A common medical practice is to treat single-pathogen infections with broad-spectrum antibiotics that affect several different bacterial species, rather than use a drug that targets particular pathogens. Comment on this practice from the perspective of an evolutionary biologist.**

   Broad-spectrum antibiotics that affect many different species of bacteria, nonpathogens as well as pathogens, tend to increase the number of different bacterial populations that are simultaneously exposed to an agent that selects for resistance.

3. **Can the emergence of antibiotic resistance be avoided altogether? Explain.**

   Unlikely. In principle, any targeted aspect of pathogen biology is subject to

mutation, and some of these mutations will promote resistance. You would need to find just the right combination of a drug that targeted a particularly inflexible aspect of pathogen metabolism. Considering that the critical and highly conserved ribosomal RNA can experience mutations that make it resistant yet still functional, the hope that resistance can ultimately be avoided is diminished.

Exercise 4.3: Observing Evolution in Action

In Activity 3, you explored some experimental approaches, such a directed evolution, that allow researchers to manipulate the evolution of populations of RNA molecules in the laboratory. In this exercise, you apply selection to shape the evolution of a population of living cells.

You will observe the effect of varying concentrations of streptomycin on a population of the bacterium *Escherichia coli* (*E. coli*). You will accomplish this by growing the bacterium in the presence of a concentration gradient of the antibiotic. You can easily establish a gradient on a single petri dish using the gradient plate technique (Figure 4.2). Across time, diffusion of the streptomycin from the upper layer into the lower layer establishes a concentration gradient of the drug (from low to high), running from one side of the plate to the other. Cells growing at different locations on the plate are thus subjected to different concentrations of the drug.

After the bacterial population has grown in the streptomycin gradient, you will observe the distribution of growth across the plate and draw conclusions about the population.

Materials

- **1 water bath (for the entire class)**
- **1 sterile petri dish**
- **1 automatic pipette (0.1–0.2 ml)**
- **1 bent glass rod "hockey stick"**
- **beaker of 70% ethanol**
- **marking pencil**
- **1 tube of starting bacterial population (a 24-hour culture of *E. coli* grown in nutrient broth)**
- **2 10-ml tubes of trypticase soy agar**
- **1 tube of stock streptomycin solution (10 mg per 100 ml water)**

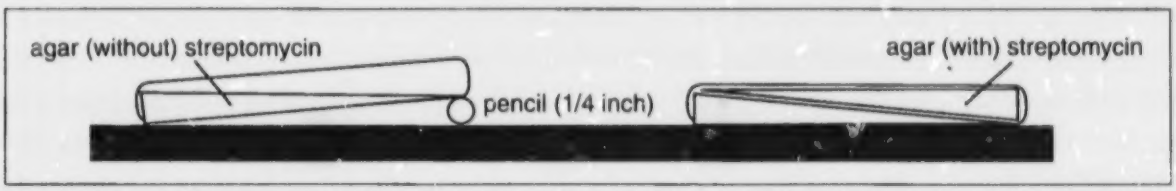


Figure 4.2 Preparation of an antibiotic gradient plate. (1) Pour the lower layer of nutrient agar, containing no antibiotic, at an angle and allow it to solidify. (2) Pour the upper layer of nutrient agar, containing 1 µg/ml of streptomycin, while the plate is level.

Procedure

1. Melt the trypticase soy agar in both tubes by immersing them in the hot water bath. When the agar has melted, place the tubes at 45°C.

2. Use the marking pencil to draw an arrow across the middle of the petri dish bottom (on its *outside* surface).

3. Place a pencil under the edge of the plate perpendicular to the arrowhead. Pour in enough molten agar to just cover the entire bottom surface. Allow the agar to solidify in the slanted position.

4. While the first layer is solidifying, use a sterile pipette to add 0.1 ml of the stock streptomycin solution to the second tube of molten agar. Mix well by vortexing or swirling the tube.

5. Place the petri dish in a level position and pour in enough of the streptomycin-containing agar to just cover the high edge of the lower agar layer. Allow to solidify.

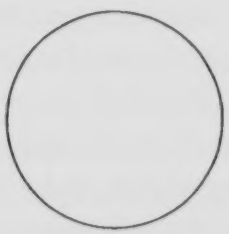
6. With a sterile pipette, deposit a 0.2 ml inoculum of the *E. coli* population on the surface of the agar. Use an alcohol-dipped and flamed bent glass rod to spread the 200 million or so individual cells uniformly over the entire surface of the agar.

7. Label the outside of your plate and incubate in an inverted position for 48 hours at 37°C.

8. Following the first incubation, examine the distribution of bacteria on the plate. You will see areas of confluent growth, where many individuals from the original inoculum are growing crowded together, and areas where discrete colonies are separated by regions showing no growth. Recall that each colony consists entirely of the offspring of a single original cell at that location on the plate, all dividing together. (In a genetic sense, you can think of each colony, with its approximately 10^7 cells, as the clonal equivalent of one cell from the original inoculum).

   a. On the following diagram, draw your arrow and indicate the *high* and *low* ends of the streptomycin gradient.

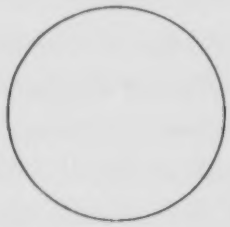
   b. Indicate on the diagram the locations of confluent growth and of discrete colonies.



Then select two or three isolated colonies in the middle region of the plate and, using a sterile inoculating loop, streak each colony toward the *high* streptomycin side of the plate.

9. **Reincubate the plate in an inverted position for an additional 48 hours at 37°C.**

10. **Following the second incubation, indicate on the following diagram the pattern of growth that resulted from the colonies that you streaked toward the high streptomycin concentration.**



Challenge Questions

1. **What can you conclude about those cells in your original *E. coli* population that were able to grow in the higher concentrations of streptomycin, even though they had never been exposed to the drug before you plated them? How does this result relate to the genetic diversity that existed within the original population of bacteria used in the experiment?**

   The cells able to grow at the higher streptomycin concentrations already possessed a degree of resistance to the drug. These pre-existing resistance mutants reflect genetic diversity that arises within any population by random mutations occurring during nucleic acid replication.

2. **Assume that you take a few cells from one of the colonies growing at high streptomycin concentration and grow a population of cells from them. How would this population differ from the original one in terms of its proportion of streptomycin-resistant individuals? How does the emergence of this new population represent the workings of selection and evolution?**

   The new population would consist almost entirely of resistant individuals, because they are all descendants of resistant individuals. A very small fraction of these new cells might be sensitive once again, if they happened to undergo a reversion mutation. The emergence of this new population of resistant cells is a classic example of the workings of the three basic processes of evolution: replication, genetic variation, and selection.

3. **In your experiment, the levels of streptomycin obviously were not high enough to kill all the bacteria in the population. State the parallels between the results of your experiment and the emergence of resistance among the infecting bacteria in patients treated with inadequate doses of antibiotic.**

   At low, suboptimal doses, antibiotics act as selective agents rather than as effective killing agents. Patients (or domestic animals) exposed to such ineffective doses likewise become unintentional "experiments" through which evolution can produce resistant strains.

4. **Recall that streptomycin inhibits bacterial growth by binding to the organism's 16S ribosomal RNA. Bacteria become resistant to streptomycin primarily as a result of mutations in the 16S rRNA or in one of the ribosomal proteins normally associated with this RNA. Can you determine from the results of this experiment whether the resistance in your bacterial population is due to a change in the rRNA or in the ribosomal protein? Review Figure 4.1 and try to think of an answer that uses the technique of nucleic acid sequencing.**

It is not possible to know from these results alone whether the resistance is due to change in the rRNA gene, the protein-encoding gene(s), or even possibly in genes that code for enzymes involved in the metabolism and breakdown of streptomycin. Sequencing the relevant genes (or using restriction mapping) would allow you to pinpoint the mutations responsible (assuming that a sufficient database of resistance-causing mutations exists, as is the case for the 16S rRNA).

FOR YOUR INFORMATION
The Evolution of RNA Viruses

Many viruses have genomes of RNA. Some are important human pathogens that have significant public health and economic impacts: common cold, flu, human immunodeficiency virus (HIV), and measles, among others. Many examples are also found among plant viruses, several of which have major ecological and economic impacts in agriculture and forestry.

As in the case of bacteria, any virus mutates as it replicates. A distinctive feature of RNA viruses, however, is their very high rate of mutation (10^{-3}–10^{-4} per base pair per replication, a rate 100,000 times greater than for cells). As discussed in Exercise 4.1, the frequency with which mutations arise in any population is determined in part by how often the organisms reproduce. Viruses replicate millions of times each day, so random mutations are constantly arising.

Another factor contributing to high mutation rates of viruses is the relative infidelity of viral nucleic acid replication. The replicase enzyme that copies RNA genomes occasionally makes random errors, inserting the incorrect monomer (for example, A opposite G, or U opposite C). The polymerases that copy DNA also make errors, but unlike RNA viruses, cells have evolved a molecular quality-control, proofreading mechanism that is able to correct mistakes most of the time and thus keep DNA's mutation rate low. The relative "sloppiness" of RNA replication enhances the rate of new mutations. It has been estimated that among the HIV viral population in a single infected person, each of the virus's 10,000 RNA bases is mutated more than 10,000 times *each day*.

As a result of their high mutation rates, any RNA virus population will contain a high level of sequence variation. Thus the "genome" of a population of RNA viruses is not a single unique sequence, but rather a population of many related variants. Because mutations occur randomly, some viral genomes might have escaped mutation entirely, while others may have many mutations. Most of the mutations are harmful, and those viruses will not survive. Some genome changes, however, will by chance confer an advantage for viral propagation or other aspects of viral behavior. Because the evolutionary principle of selection applies to viruses as well as to free-living organisms like bacteria, we can make certain predictions about the general course of virus evolution in the face of medical attempts to thwart them with antiviral drugs.

Exercise 4.4: Evolution in a Microcosm

In this exercise, you will see how an RNA virus such as HIV can evolve even within the microcosm of a single infected individual.

Figure 4.3 lists the variant strains of HIV that emerged and predominated at different times during the course of a particular patient's treatment regimen. Nucleotide base sequences were determined in samples of viral RNA (or proviral DNA) taken from the patient at various times before, during, and after treatment with the antiviral drug zidovudine (AZT) alone. AZT is a commonly used component of antiviral drug cocktails because it can inhibit reverse transcriptase (RT). RT is an important, virally encoded enzyme responsible for replicating the virus's RNA and copying it into a DNA provirus form. The DNA provirus can then integrate into the infected cell's genome to produce a state of chronic infection.

Each of these viral variants was found to contain one or more mutations in the RNA gene that encodes reverse transcriptase. Rather than impairing the enzyme's function, these particular mutations rendered the enzyme less susceptible to the AZT drug, thus conferring a degree of drug resistance on the variant virus. Variants that have more than one such mutation in the same RT gene are designated with multiple numbers (for example, variant *70;215* has two different mutations). Each number in the designation refers to a mutation in RNA that affects the amino acid at that position number in the RT protein. Variant *70;215*, for example, has two mutations in its RNA, one affecting amino acid 70 and the other affecting amino acid 215 of the RT protein.

You can see in Figure 4.3 that the virus population changed across time in this patient. The change in the viral population is a good example of evolution in a microcosm—in many ways similar to the test tube evolution of RNA that you explored in Activity 3.

You can better understand the historical pathway of HIV evolution within this patient by identifying the ancestor-descendant relationships among the different variant strains. Figure 4.4 is a branching diagram (a *phylogeny*) that allows you to summarize these relationships.

	Time of Sample During Treatment (Weeks)						
	Before Treatment	17	22	56	81	110	After Treatment, 148
Variant Strain Present	*pre*	*70*	*70*	*70;215*	*70;215;41*	*215;41*	*215;41*

Figure 4.3 HIV variant strains appearing during the course of infection and treatment.

Procedure

1. **Use the data in Figure 4.3 to fill in the phylogeny diagram. In the square brackets [ ], write in the variant strain that was present at a particular time (for example, *215;41*).**

2. **In the parentheses ( ), write the designation number of the particular new mutation (for example, *215*) that arose in the population at the indicated point in time.**

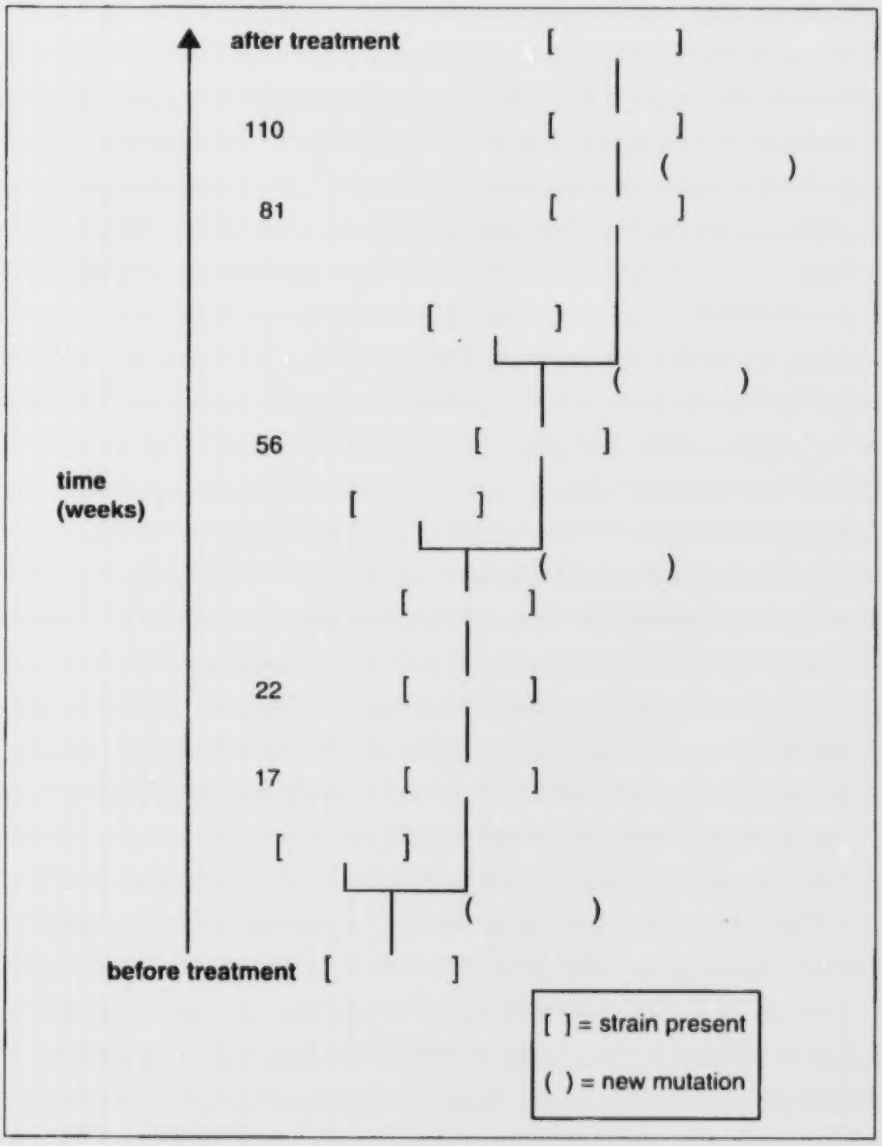


Figure 4.4 A phylogeny of ancestor-descendant relationships showing an evolution of HIV in a single patient.

Challenge Questions

1. **The viral population in this patient changed across time. Explain this change in evolutionary terms. What role did AZT play in shaping the course of this change?**

   Evolution occurred due to the combined action of replication, random variation, and selection. AZT played the role of a selective agent that directed this evolutionary process.

2. **What do you think happened in the branches (lineages) that appear to terminate during the course of the infection?**

   These strains were less resistant to AZT and thus less competent for replication than later-emerging strains. These strains eventually became "extinct" within the patient due to the combined effects of the drug and competition by the more resistant strains.

3. **Which combination of mutations appears to confer the most AZT resistance on the virus during the course of this study? Explain the reason for your choice.**

   215;41. This strain persists and replaces the others.

4. **What mutational mechanism can account for the eventual loss of the *70* mutation in the surviving virus lineage?**

   The mechanism is random reverse mutation (*reversion*) of the *70* mutation to its original nucleotide base.

5. **The emergence of strain *215;41* sometime between weeks 81 and 110 of therapy, and its persistence after the end of therapy, implies that the rate of change in the viral population slowed when therapy stopped. Explain this slowing in evolutionary terms. What might you expect to happen if therapy with AZT alone were reinstituted? What can you conclude about the long-term effectiveness of treatment with only a single drug?**

   Removal of the drug removed a selection pressure that favored and sped up the establishment of new strains. Reintroducing the drug could accelerate the virus's evolution again. Single-drug treatment for rapidly evolving viruses such as HIV is unlikely to be effective in eradicating the virus.

6. **After reading the FYI essays titled *The Evolution of RNA Viruses* and *Combating Viral Disease*, answer the following questions:**

   a. **Antiviral treatment protocols commonly use concurrent administration of two or more antiviral drugs, with each drug targeting a different essential viral function or gene product. Explain the evolutionary rationale for this combination treatment.**

The combination therapy approach greatly reduces the likelihood that a resistant strain of virus will emerge. The probability that a viral lineage will randomly mutate at three separate genes and acquire resistance to three independent drugs is much smaller than the probability of resistance to any one of the drugs.

b. Explain the evolutionary rationale for starting treatment with antiviral drugs as soon as possible following exposure to the virus.

Viruses replicate and mutate at very rapid rates. Drug treatment is started early, before the genetic diversity of the initial viral population has a chance to increase through random mutations. This approach minimizes the likelihood that drug-resistant mutants will have had a chance to arise spontaneously before treatment has begun.

FOR YOUR INFORMATION
Combating Viral Disease

Antibiotics kill bacteria and other cellular microbes by interfering with cellular structure or function. Because cells are complex structures, there are many points at which they can be vulnerable to different antibiotics. You have seen how the aminoglycosides interfere specifically with bacterial protein synthesis by binding to bacterial rRNA. Penicillin, in contrast, impairs synthesis of the bacterial cell wall. (Eukaryotic cells are unaffected because they lack such walls.) Still other classes of antibiotic interfere with functions of the bacterial cell membrane, such as ion transport.

Because viruses are noncellular structures (they must parasitize living cells to reproduce), they are not affected by antibiotics. Taking an antibiotic for your head cold or flu is not only a waste of time and money, it needlessly exposes your normal bacterial flora to a selective pressure that promotes antibiotic resistance. A variety of other chemical agents known as antivirals have been developed that target different steps in the viral reproduction cycle. Antivirals typically interfere either with the entry of the virus into the cell, the machinery of viral genome replication, or the assembly of progeny virus particles. In the case of HIV 1, the viral cause of AIDS, antivirals in current use fall into two main categories: those that interfere with key viral enzymes needed for viral RNA replication (for example, inhibitors of reverse transcriptase or integrase), and those that impair the maturation of viral proteins and the assembly of progeny virus particles (for example, protease inhibitors).

The available antiviral agents have not yet been able to cure AIDS. However, recent treatment protocols that are based on evolutionary principles have shown considerable benefit in reducing the number of virus particles in HIV-infected individuals and in delaying the onset of immune system collapse and full-blown AIDS. These evolution-based protocols share the following features:

- **They employ two or more antivirals administered concurrently.**
- **Each antiviral in the mix targets a different viral function or gene product (for example, an RT inhibitor combined with a protease inhibitor).**
- **Treatment is started as soon as possible following initial exposure to the virus.**

Exercise 4.5: Evolution in a Larger Population

You have seen how, even in the microcosm of a single patient, viruses can evolve important new properties such as drug resistance in response to "environmental" factors. Here you will examine the relatedness among several strains drawn from the wider population of HIV viruses that infect humans worldwide.

The following partial nucleotide sequence runs from base 2726 to base 2750 of the HIV gene encoding reverse transcriptase, the viral replicating enzyme dealt with in Exercise 4.4. You can view the entire 10,000+ bases of the HIV genome, along with its known mutations, by checking out the HIV Sequence Database on the Internet at http://hiv-web.lanl.gov/.

base 2726 base 2750
...CCATAAAGAAAAAAGACAGTACTAA...

Note that the bases shown are the proviral DNA form of the virus isolated from the human genome.

The reverse transcriptase gene has been sequenced for several different strains of HIV from several different patients worldwide. Many mutations have been documented in the gene, and those of particular interest are ones that confer enhanced virulence and/or drug resistance to the virus.

Figure 4.5 focuses on particular base positions within the above sequence, and all are known positions of mutations that confer resistance to the antiviral drug AZT. The six different mutant sequences shown were isolated from six different patients who had undergone therapy with AZT. A "wild type" sequence, from an HIV strain that had not been exposed to AZT, is also shown.

Sequence		Base Position					
		2726	2735	2740	2746	2749	2750
		C	A	G	A	A	A
wild type	1	T	A	G	A	A	A
mutant sequence	2	C	A	G	G	A	A
mutant sequence	3	T	G	A	A	A	A
mutant sequence	4	C	A	G	G	G	G
mutant sequence	5	T	G	G	A	A	A
mutant sequence	6	C	A	G	G	G	A

*Read sequences from left to right.

Figure 4.5 Wild type and mutant bases.

Procedure

Arrange the sequences in Figure 4.5 into the phylogeny shown in Figure 4.6, which depicts a plausible ancestor-descendant relationship among these sequences. To do this, you will use the method of parsimony. Parsimony is a decision-making approach in the field of taxonomy that applies the following assumption: *Those sequences that are most closely related to one another will differ by the smallest number of base changes.* For example, sequences that differ from each other at only one base position are presumed to be more closely related than sequences that differ at more than one position.

1. **Consider which of the sequences is likely to be the most ancestral. Write this sequence in the square brackets [ ] at the bottom of the phylogeny.**

2. **Examine the other sequences and apply the parsimony principle to find the "family" relationships that emerged from this ancestral strain. Fill in the appropriate sequences in the square brackets [ ] at the top of the phylogeny.**

3. **In the parentheses ( ), fill in the specific base-change mutation that occurred and caused divergence of sequences in that branch of the phylogeny (for example, C to T).**

Challenge Questions

1. **As you can see from Figure 4.6, two major lineages (branches) appear to have diverged early from the ancestral sequence. Which newly acquired base at which position is shared by all members of the major lineage on the left? on the right?**

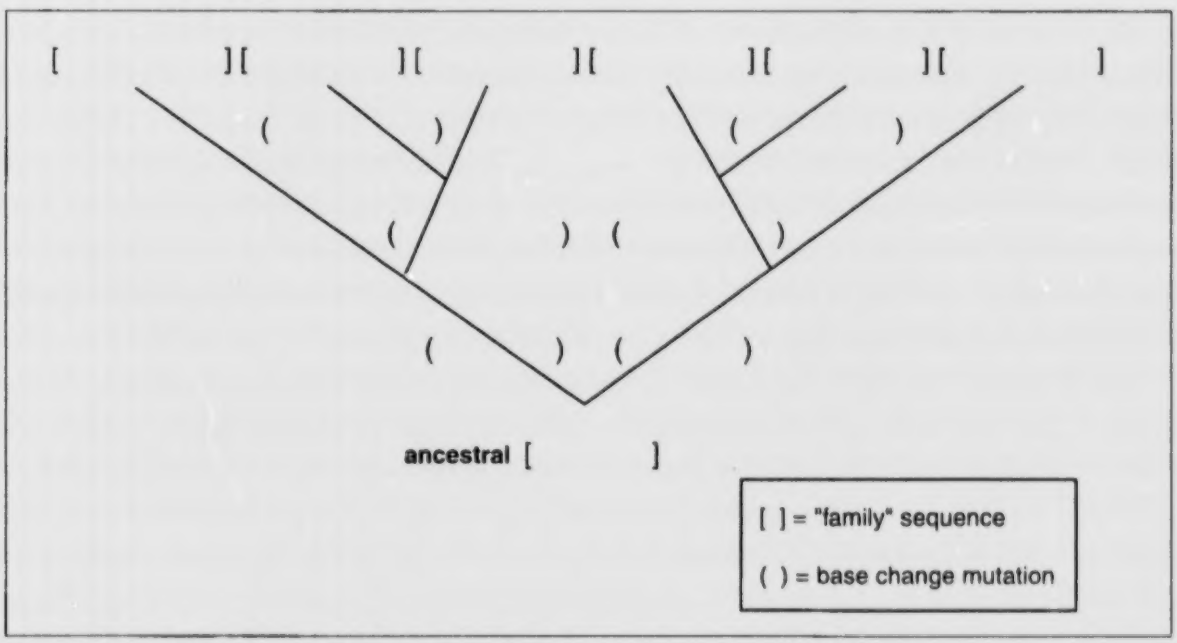


Figure 4.6 A possible phylogeny of ancestor-descendant relationships among several HIV strains.

The new base on the left is T at 2726; on the right is G at 2746.

2. **Which two sequences are more closely related to the ancestral sequence? Which two are more divergent from the ancestral? On what basis do you make these inf rences?**

   TAGAAA and CAGGAA differ from the ancestral by only one base each and are therefore more closely related to it than the others. TGAAAA and CAGGGG differ from the ancestral by two bases each and are therefore more divergent from it than the others. Inferences should be based on the parsimony principle.

3. **Recall that these mutations are in the gene encoding the viral reverse transcriptase enzyme, and each of them confers a degree of resistance to the antiviral drug AZT.**

   a. Which of the viral strains in your phylogeny have two or more resistance mutations in the reverse transcriptase gene?

   The viral strains that are double mutants are TGGAAA and CAGGGA. Those that are triple mutants are TGAAAA and CAGGGG.

   b. Based on your results from Exercise 4.4, what might you predict about the relative drug resistance of the multiple- versus the single-mutant strains of the virus?

   Multiple-mutant strains frequently show greater resistance. However, it is possible that different resist nce mutations, when present together, may undermine each other's resistance effect, leading to decreased resistance relative to the single mutants.

FOR YOUR INFORMATION
The Continuity of Life

Only one-half billion years after the earth formed, early cells similar to modern cyanobacteria were living in large, domelike colonies. Today these colonies exist in fossilized form as stromatolites. These oldest of solid clues to the ancient origins of life show us that as long ago as 3.5 billion years, cell structure and biochemistry quite similar to modern life already existed. But how do we observe evidence of the world that existed before 3.5 billion years ago? This was a time before cells, a time when some type of ancient molecule was first carrying out essential activities, such a replication, that characterize life.

Among our best evidence for events on the ancient earth are fossils. But molecules do not make good fossils; they are too small and too fragile. How, then, can we examine that early world? Fortunately, the molecules of our modern world offer a glimpse of the shadows of past events. This connection exists because life involves continuity: Despite great changes that occurred accross time, life and its molecules have existed continuously since its origin almost 4 billion years ago from a single common ancestor. We know this is the case because the molecular machinery of all living cells—from bread mold to bald eagles—is fundamentally similar.

FOR YOUR INFORMATION (continued)

When the base sequences of ribosomal RNAs from many different organisms are compared, for example, the sequences are found to be remarkably similar, although not identical. The similarities immediately suggest that all organisms are related and derive from a single common ancestor. The differences presumably reflect the fact that mutations are constantly arising at random across time in DNA and RNA. Only a few mutations have had a chance to occur in the short time since the organisms diverged from their common ancestor. As a result, closely related organisms have closely related ribosomal RNA sequences, while distantly related organisms have correspondingly more divergent ribosomal RNA sequences. Knowing the degree of relationship between organisms makes it possible to construct a "tree of life" or phylogeny showing the lines of descent from one species to the next.

One of the great surprises of this phylogenetic analysis was Carl Woese's discovery that all living organisms can be divided into three great domains. These are the *eukaryotes* (cells with nuclei), the *eubacteria* (true bacteria), and the *archaebacteria* (ancient bacteria that in some respects appear to be more closely related to eukaryotes than to true bacteria). Both groups of bacteria lack nuclei and are also referred to as *prokaryotes*. Remarkable conservation of ribosomal RNA sequence is found among organisms as diverse as the bacterium *Escherichia coli* (which populates our intestines), the archaebacterium *Halobacterium halobium* (which colors the salt flats of San Francisco Bay red), and the unicellular eukaryote *Trypanosoma brucei* (the cause of African sleeping sickness). This molecular similarity strongly supports the idea of an underlying relatedness among these organisms. The fact that ribosomal RNA has changed relatively little across time supports the idea that this molecule plays a key role in protein synthesis, almost certainly as the catalyst of peptide bond formation. (RNA's ability to carry out this reaction also reinforces the notion that the very first ribosomes consisted simply of a catalytic RNA able to join amino acids, and that modern ribosomal RNA has retained these catalytic functions despite assistance from more than 50 ribosomal proteins added later). Recall that certain fungal-derived antibiotics such as streptomycin and neomycin exploit both the indispensability of rRNA and the subtle sequence differences between prokaryotic and eukaryotic rRNAs to target and kill bacteria (recall Exercises 4.2 and 4.3).

Biologist A.G. Cairns-Smith used the following analogy to describe the continuity of modern and ancient forms of life: "None of the fibers in a rope has to stretch from one end to the other, so long as they are sufficiently intertwined to hold together sideways." For life, the genetic information passed repeatedly from one generation to another acts like the fibers in a rope and provides continuity across time. No single gene sequence stretches unaltered all the way across the billions of years, but life's collection of continuously replicated, related genetic information does form an unbroken chain from the early earth to the present.

Extension Exercise—Exercise 4.6: Comparing RNA Sequences

This exercise is a useful demonstration of how the comparison of RNA sequences can be informative in establishing phylogenetic relationships among organisms.

The following partial sequences consist of aligned regions of the small subunit rRNA from six different organisms: two archaebacteria, two eubacteria, and two eukaryotes. Cut out each sequence and paste the sequences onto separate index cards. Distribute a set of the six cards to each group of students. Without telling them that the sequences represent the three domains of the Woese classification, have the students examine the sequences by eye and attempt to arrange them into subgroups of the most closely related sequences.

sequence 1 GC-ACG- - - -AAA- -GUG-CGA- -C-G-GGG- - - - - - - -GNA-UC-CCA-A-GU

sequence 2 GC-GCG- - - -AAA- -GCG-CGA- -C-G-GGG- - - - - - - -GNA-CC-CCA-A-GU

sequence 3 CC-UGA- - - -CAC- - -GG-GGA- -G-G-UAG- - - - - - - -UGA-CA-AUA-A-AU

sequence 4 CC-UGG- - - -CAC- - -GG-GGA- -G-G-UAG- - - - - - - -UGA-CG-AAA-A-AU

sequence 5 GG-GGG- - - -AGA- -CCC-UGA- -C-G-CAG- - - - - - - -CAA-CG-CCG-C-GU

sequence 6 GG-GCG- - - -AAA- -GCC-UGA- -C-G-GAG- - - - - - - -CGA-CA-CCG-C-GU

- Sequences 1 and 2 (*Methanococcus deltae2* and *Methanococcus igneus*) are both from archaebacteria.
- Sequences 3 and 4 (the plant *Arabidopsis thaliana* and the animal *Mytilus californianus*) are both from eukaryotes.
- Sequences 5 and 6 (*Clostridium botulinum* and *Treponema pallidum*) are both from eubacteria.

The students will initially be struck by the apparent similarity of all these sequences. On closer inspection, however, they should easily be able to discern the three subgroups corresponding to the three-domain classification.

You can also have the students log onto the SSU rRNA Database to access rRNA sequence data on many other organisms (http://rrna.uia.ac.be/ssu/).

Copymasters

BEST COPY AVAILABLE

Copyright © 2000 by BSCS. Permission granted to reproduce for classroom use.

Copymaster for Alternate Exercise 1.2

Use these primary base sequences to prepare strips for the vinyl tubing models. Construct four different 34-base sequences from this as follows:

1. Make two photocopies of this sheet for each team of four students. Prepare enough strips so that half of your students receive a copy of sequences 1 and 3 and half receive a copy of sequences 2 and 4.
2. Note that each 34-base sequence is printed as four half-strings (two identically labeled 5' strings and two identically labeled 3' strings). Use a paper cutter or a utility knife and straight edge to cut out each sequence, then cut apart the two identically labeled 5' rows from the two identically labeled 3' rows.
3. Fold each pair of identically labeled rows along the dotted line and glue together.
4. Overlap the right-most base of the 5' string with the identical left-most base of the 3' string and glue the half-strings together, creating a sequence 34 bases long that is labeled on both sides.

Seq. 1	5'-U	A	G	A	G	U	C	U	G	A	U	C	G	A	C	A	G
Seq. 1	5'-U	A	G	A	G	U	C	U	G	A	U	C	G	A	C	A	G
G	G	U	G	C	U	C	U	A	C	C	G	A	A	C	C	U	3'-A
G	G	U	G	C	U	C	U	A	C	C	G	A	A	C	C	U	3'-A

Seq. 2	5'-C	U	A	U	A	C	C	C	G	A	A	A	A	U	C	A	G
Seq. 2	5'-C	U	A	U	A	C	C	C	G	A	A	A	A	U	C	A	G
G	G	U	G	C	U	C	U	C	C	C	G	A	A	C	C	U	3'-G
G	G	U	G	C	U	C	U	C	C	C	G	A	A	C	C	U	3'-G

Seq. 3	5'-C	G	U	C	G	G	A	A	A	C	G	U	A	U	C	C	U
Seq. 3	5'-C	G	U	C	G	G	A	A	A	C	G	U	A	U	C	C	U
U	U	U	A	C	A	C	C	G	A	A	G	C	A	U	A	C	3'-G
U	U	U	A	C	A	C	C	G	A	A	G	C	A	U	A	C	3'-G

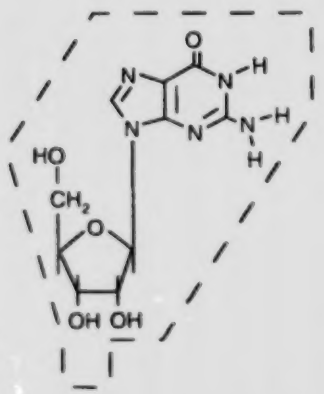
Seq. 4	5'-G	C	G	C	G	A	U	A	C	U	C	A	A	C	C	C	C
Seq. 4	5'-G	C	G	C	G	A	U	A	C	U	C	A	A	C	C	C	C
C	A	A	A	A	U	U	G	C	C	A	G	A	A	G	C	G	3'-C
C	A	A	A	A	U	U	G	C	C	A	G	A	A	G	C	G	3'-C

BEST COPY AVAILABLE

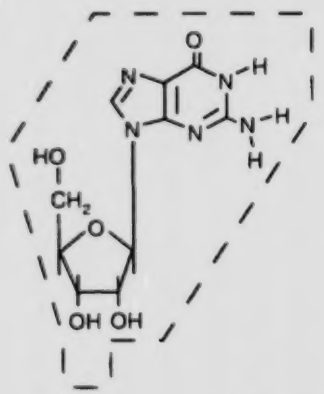
Copyright © 2000 by BSCS. Permission granted to reproduce for classroom use.

Copymaster for Exercise 2.1 b

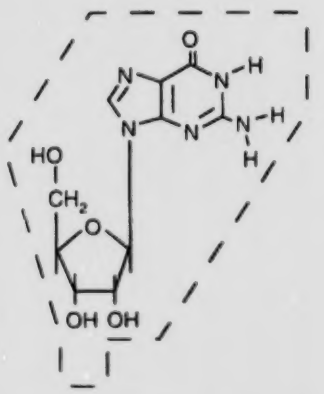
Create paper models of the guanosine molecule and OH group by photocopying this page and cutting out the molecules. Make enough copies so that each team has one guanosine cutout and one OH (hydroxyl) group.



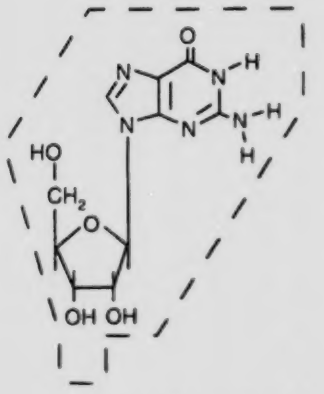
guanosine molecule



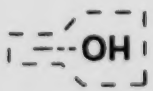
guanosine molecule



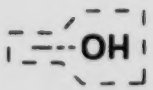
guanosine molecule



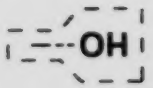
guanosine molecule



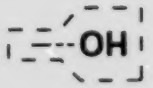
OH group



OH group



OH group

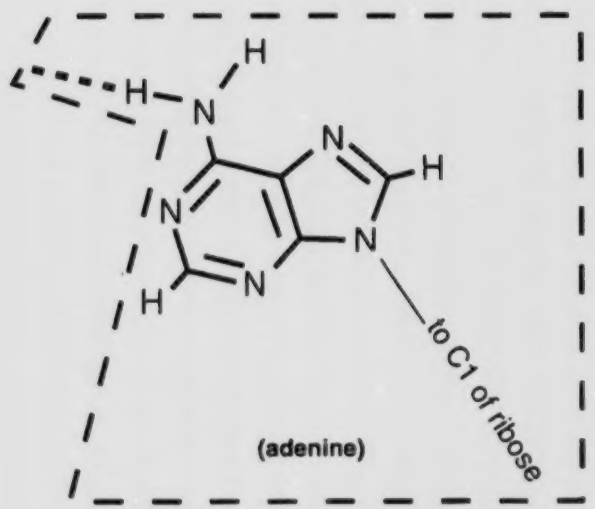


OH group

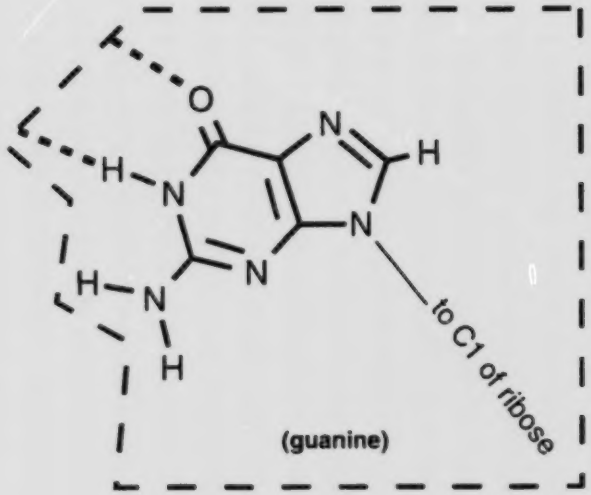
Copyright © 2000 by BSCS. Permission granted to reproduce for classroom use.

Copymaster for Exercise 3.1

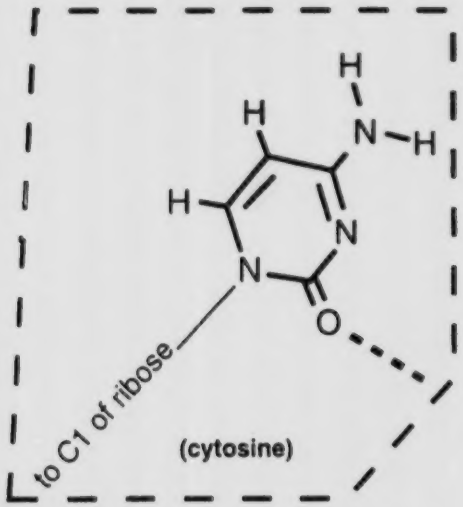
Create paper models of the bases by photocopying this page and cutting out the molecules. Make enough copies so that each team has one cutout of each molecule.



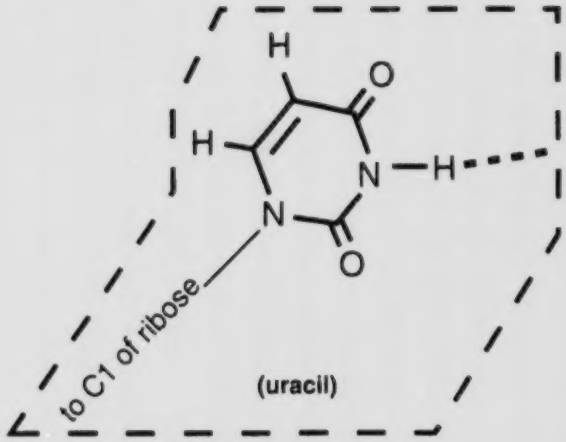
(adenine)



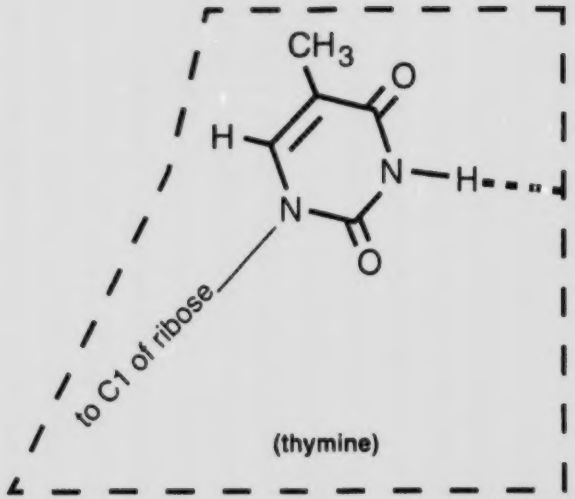
(guanine)



(cytosine)

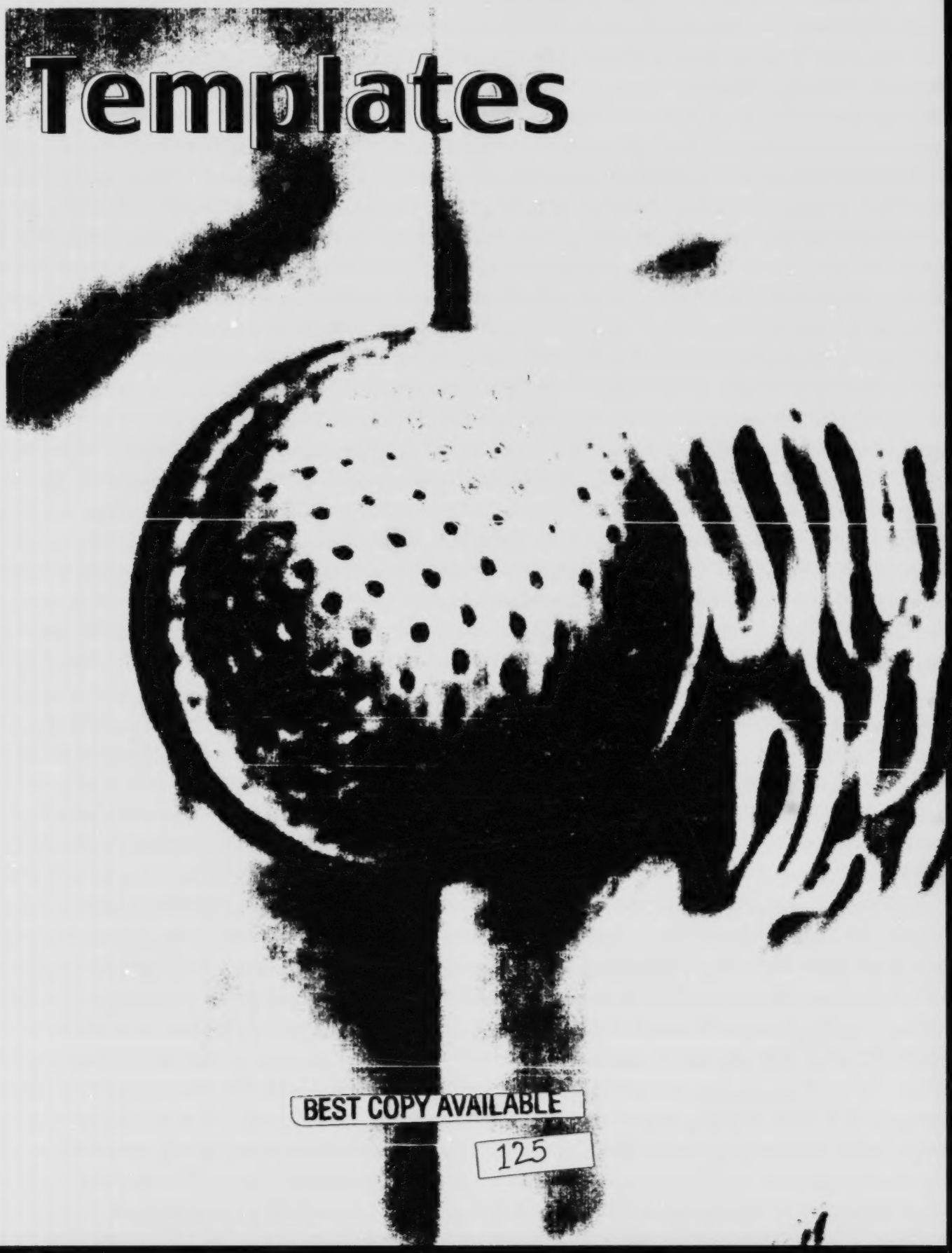


(uracil)



(thymine)

Templates



BEST COPY AVAILABLE

Copyright © 2000 by BSCS. Permission granted to reproduce for classroom use.

Left Template for Exercise 1.2

To create the template for color-coding, lay the left and right templates next to each other and align the sequences. Do not overlap the bases. Tape the pages together.

To color-code the pipe cleaner models, lay two 50.5-cm pipe cleaners on a sequence. Tape the pipe cleaners in place with removable tape. Color each base with the appropriate color: yellow = adenine (A), green = uracil (U), black = guanine (G), and white = cytosine (C). Be sure to color completely around the pipe cleaner. Double-check that you have colored the pipe cleaners correctly before proceeding.

sequence 1

5'-U	A	G	A	G	U	C	U	G	A	U	C	G	A	C	A	G

sequence 2

5'-A	U	A	U	A	C	C	C	G	A	A	A	A	U	C	A	G

sequence 3

5'-C	G	U	C	G	G	A	A	A	C	G	U	A	U	C	C	U

sequence 4

5'-G	C	G	C	G	A	U	A	C	U	C	A	A	C	C	C	C

Copyright © 2000 by BSCS. Permission granted to reproduce for classroom use.

Right Template for Exercise 1.2

sequence 1

G	U	G	C	U	C	U	A	C	C	G	A	A	C	C	U	3'-G

sequence 2

G	U	G	C	U	C	U	C	C	C	G	A	A	C	C	U	3'-G

sequence 3

U	U	A	C	A	C	C	G	A	A	G	U	A	U	A	C	3'-G

sequence 4

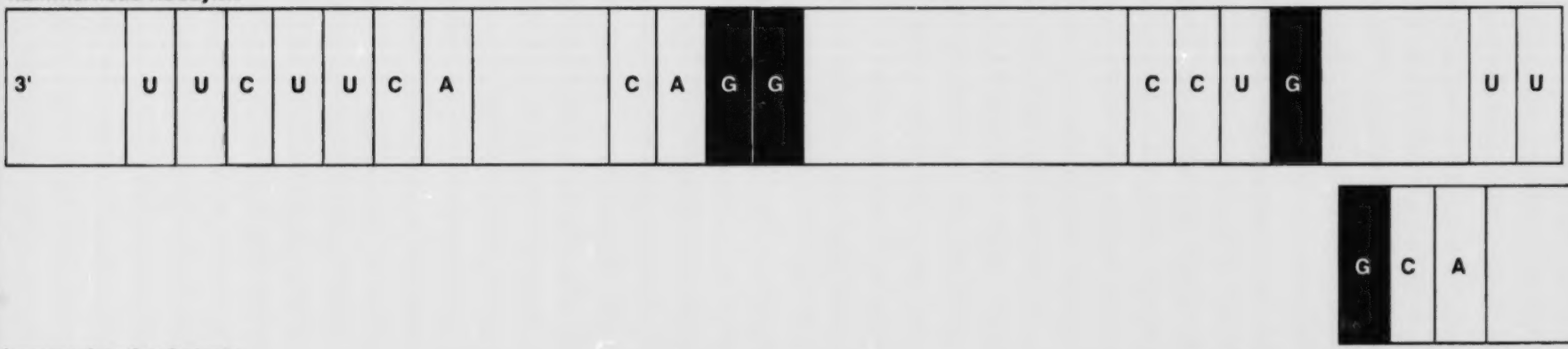
A	A	A	A	U	U	G	C	C	A	G	A	A	G	C	G	3'-C

Copyright © 2000 by BSCS. Permission granted to reproduce for classroom use.

Template for Exercise 2.1

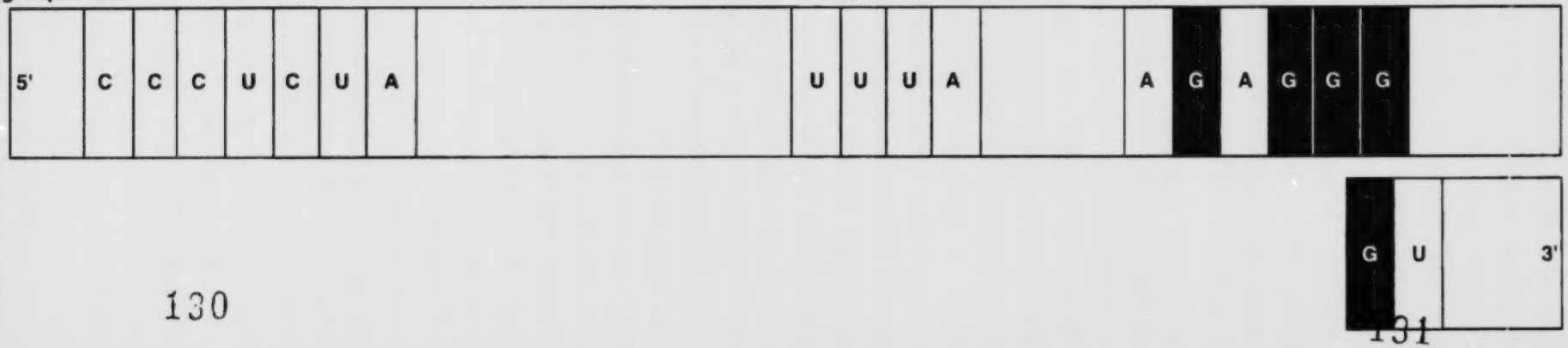
To color-code the pipe cleaner models, lay a 27-cm pipe cleaner on a sequence. Align the left end of the pipe cleaner with the left (5') end of the sequence. Tape the pipe cleaner in place with removable tape. Working from left to right, color each base with the appropriate color: yellow = adenine (A), green = uracil (U), black = guanine (G), and white = cytosine (C). Be sure to color each base according to the width shown and to color completely around the pipe cleaner. To finish, reposition the pipe cleaner so that the right end aligns with the right (3') end of the sequence. Color the remaining bases. Double-check that you have colored the pipe cleaners correctly before proceeding.

hammerhead ribozyme



hammerhead substrate

group I intron



Student Pages

BEST COPY AVAILABLE

Activity 1
RNA Structure: Tapes to Shapes

Introduction

What does the shape of a thing have to do with the job it performs? You no doubt can think of a variety of familiar household objects that clearly show the relationship between shape and function, such as a metric wrench that fits only metric-sized bolts or a key that must have exactly the right shape to unlock a lock. Molecules also display an essential relationship between structure and function.

This activity offers you a combination of hands-on exercises and short essays that examine structure and function in RNA. To discover for yourself the answers to the Challenge Questions posed in an exercise, you should perform the hands-on inquiry portions of the exercise first. Then read the accompanying For Your Information (FYI) essays to help you understand your findings. In Activity 1, you will construct models of different RNA molecules to investigate the following:

- the principles underlying shape determination in RNA,
- the variety of shapes in RNA,
- how shape can influence function in RNA,
- how alterations (mutations) in RNA building blocks can influence structure and function, and
- how comparisons among related molecules from different species can be useful.

Exercise 1.1: RNA Sequences: Primary Structure

Procedure

Read the following text, then answer the Challenge Questions.

We can use the analogy of an elongated tape as a good starting point for exploring the complex shapes of RNA molecules. RNA molecules are relatively large *polymers*, molecules composed of many smaller *monomer* subunits; the tape is analogous to the *linear sequence* of building blocks that make up an RNA molecule. You may already be familiar with proteins as biological polymers; their primary structure consists of long chains of covalently linked amino acid monomers.

BEST COPY AVAILABLE

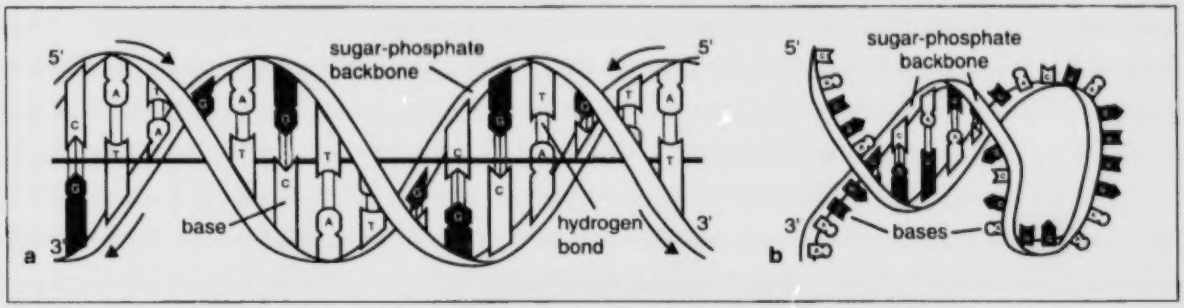


Figure 1.1 The structure of nucleic acids. (a) In DNA, alternating sugar and phosphate molecules form a linear backbone joining nucleotide subunits into chains. Two chains are held together by specific base pairing and coiled around a central axis to form a double helix; the strands run in opposite, antiparallel directions. (b) A single linear RNA strand folds by undergoing specific base-pairing interactions (A to U and G to C) within the strand.

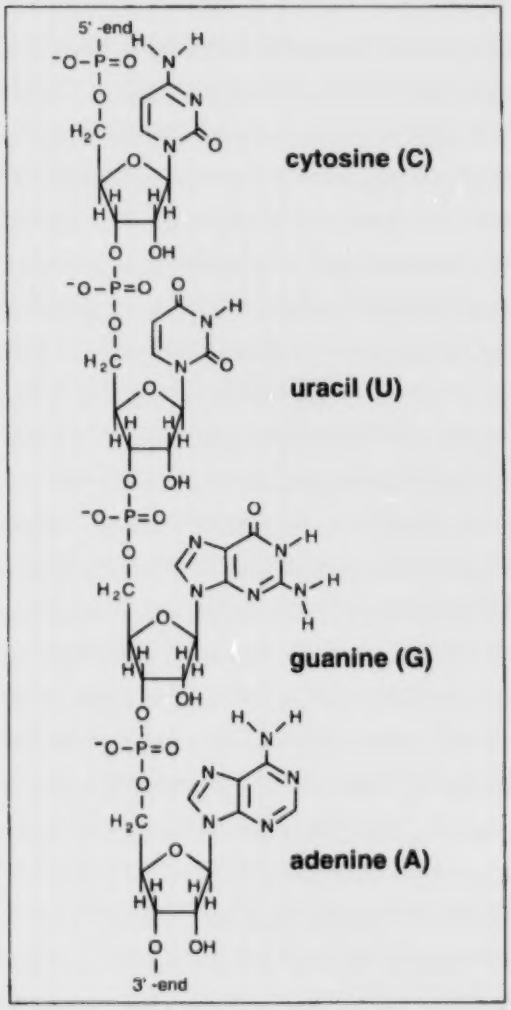


Figure 1.2 The primary sequence of a short section of single-stranded RNA illustrates the chemical structure and linkage of the monomer nucleotides.

In the case of nucleic acid polymers, RNA and DNA, the monomer subunits are called *nucleotides* and are linked together covalently into a long chain or tapelike structure (Figure 1.1). Different RNA molecules may be made up of tens to thousands of nucleotide monomers. The nucleotides themselves are made up of smaller components, namely a sugar molecule (ribose in RNA, deoxyribose in DNA), a phosphate, and a nitrogen-containing base. The base is the "business end" of the nucleotide; it interacts with other nucleotide bases in precise ways. Four different types of nucleotide occur in RNA, each having a different base: adenine (A), guanine (G), cytosine (C), and uracil (U). (In DNA, the base thymine (T) replaces uracil (U). Scientists typically refer to the individual nucleotide monomers of RNA and DNA simply as bases.

The linear order of bases in a nucleic acid is referred to as its *primary structure* (1° structure) (Figure 1.2). Recently developed biochemical methods for sequencing nucleic acids make it possible to determine in just hours the linear sequences of nucleotides in DNA or RNA that would have taken weeks to determine 15 years ago. This and several other technological breakthroughs have made possible the Human Genome Project, a worldwide effort to sequence completely the genetic material of *Homo sapiens*, as well as that of several other species. To date, the genomes of 25 bacterial species, yeast, the nematode worm *C. elegans*, and the fruit fly *Drosophila* have been sequenced completely.

Consider the nucleotide base sequence data for the four different RNA molecules shown in Figure 1.3:

Seq. 1	5′	U	A	G	A	G	U	C	U	G	A	U	C	G	A	C	A	G	G	U	G	C	U	C	U	A	C	C	G	A	A	C	C	U	G	3′
Seq. 2	5′	A	U	A	U	A	C	C	C	G	A	A	A	A	U	C	A	G	G	U	G	C	U	C	U	C	C	C	G	A	A	C	C	U	G	3′
Seq. 3	5′	C	G	U	C	G	G	A	A	A	C	G	U	A	U	C	C	U	U	U	A	C	A	C	C	G	A	A	G	U	A	U	A	C	G	3′
Seq. 4	5′	G	C	G	C	G	A	U	A	C	U	C	A	A	C	C	C	C	A	A	A	A	U	U	G	C	C	A	G	A	A	G	C	G	C	3′

Figure 1.3 Primary sequences of some RNA molecules.

Challenge Questions

1. Do you see any similarities in primary sequence among any of the molecules in Figure 1.3? If so, indicate them by circling regions of sequence that appear to be conserved between molecules.

2. At this point, what predictions can you make about the relative shapes of those molecules that have conserved regions of sequence?

3. Can you mentally visualize the three-dimensional shape of each molecule?

4. Do you think it is possible for two molecules with entirely *different sequences* to have the same shape? Discuss this within your team.

Exercise 1.2: Folding the Tape: Higher Orders of Structure

As you saw in Exercise 1.1, it is not really possible to visualize a molecule's shape from a printed string of subunits. Yet this is the same type of data that a molecular biologist is confronted with after completing the biochemical sequencing of an unknown RNA molecule. Fortunately, there is a better way.

In this part of Activity 1, you will use a simple molecular model and a specific set of folding rules to discover and visualize some common structural features of RNA. You will fold simple linear models of the RNA sequences in Figure 1.3 into more complex shapes by applying folding rules.

Materials

- pipe cleaners
- colored markers
- template sheets for color-coding
- tape

Procedure

1. Each team member receives four white pipe cleaners (30.5 cm each). Take *two* pipe cleaners and make an elongated model 50.5 cm long by overlapping 11 cm of each pipe cleaner and twisting the ends together thoroughly. Repeat this procedure with your two remaining pipe cleaners to produce a second model.

2. Each team member selects a *different* sequence from Figure 1.3 and colors *two* identical pipe cleaner models of that sequence. Use colored markers and

FOR YOUR INFORMATION
Modeling in Science

In all likelihood, you were not able to envision the shape of the molecules in Figure 1.3 simply by studying their primary sequence. Large biological molecules often have complex shapes that are much too difficult to visualize from this kind of raw data or from a written formula. This is where modeling becomes necessary. Models are important tools in science, because they help us visualize structures we cannot see directly. Model building is particularly helpful in the study of biochemistry and molecular biology. You may have used molecular models in previous chemistry or biology classes to learn the structure of known molecules. Scientists also use models to discover *new* molecular shapes. Molecular models enabled Watson and Crick, for example, to discover and visualize the structure of DNA for the first time.

A variety of modeling materials have been used over the years, from the simple wire and cardboard models used by Watson and Crick for their work with DNA, to sophisticated computer-generated images (Figure 1.4). Research scientist Harry F. Noller, at the University of California—Santa Cruz, used simple pipe cleaners to begin to model some important features of RNA structure.

Figure 1.4 (a) James Watson and Francis Crick with their original wire and cardboard model of the DNA double helix. (b) RNA researcher Harry F. Noller. (c) Computer graphic models of two transfer RNA (tRNA) molecules.

Before a structural model of any molecule can be built, scientists must first gather a lot of basic data about its chemical composition and physical properties. Today new biochemical analyses, such as nucleotide sequencing, are combined with advanced methods of X-ray crystallography and spectroscopy for this purpose. Computer modeling of molecular structure enables scientists to rotate and observe virtual molecules in three dimensions, allowing rapid comparisons to known molecules. Such comparisons often provide new clues to the behavior of the molecule under study. Some modeling software allows the scientist to change atoms or molecular subunits in the model or change incubation conditions to determine the effects on the molecule's structure. Examples of computer-generated models are available on the World Wide Web. You might try the following sites:

- The RNA World (IMB-Jena) at http://www.imb-jena.de/RNA.html (an extensive site with many links to sequence and structural databases, modeling software, and books and tutorials)
- RNA Pages at http://www.kwl.t.u-tokyo.ac.jp/~ueda/RNApage.html (many links to a variety of RNA-related sites)

the following code to mark the 34 nucleotide bases of your sequence as colored patches on the models:

yellow = A green = U black = G white = C

Make each colored patch 15 mm long. To do this, lay both pipe cleaners side by side over the paper template provided by your instructor. You can now easily color both models at the same time. (It helps to tape the ends of the pipe cleaners to the template.) Be sure to color completely around each model's circumference. Double-check your colored models for coding errors before you proceed.

3. Indicate the polarity of the RNA molecule by marking the 5′-end with a black dot; the opposite end is the 3′-end. (In the real molecule, the 5′-end nucleotide would have a free phosphate group at position 5 in its ribose sugar; the 3′-end nucleotide would have a free OH group at position 3 of its ribose).

4. Distribute the models within your team as follows: two persons *each* get a copy of sequence 1 and sequence 3; two persons *each* get a copy of sequence 2 and sequence 4. Note: You can easily distinguish the different sequence models by noting that the 5′-end base is unique for each.

   Use the folding rules listed in Figure 1.5 to fold your two sequences. Hint: Work in parallel with the team member folding the same sequences as you, so that you can help each other.

5. Compare your structures with those of your teammates and with those of other teams. Reject any invalid structures that do not conform to the folding rules. The valid models represent the potential three-dimensional structures dictated by your RNA sequence.

Figure 1.5 Folding rules. Use these rules to fold linear RNA molecules into more complex shapes.

- RNA molecules are highly *flexible*. Bases *within* the molecule can pair as follows: A pairs with U; G pairs with C.
- Molecules fold so as to *maximize* the number of paired bases.
- Paired bases create a *double-stranded stem region* in the molecule; these stem regions must contain *three or more* base pairs to be stable.
- Double-stranded stem regions can only form by *antiparallel association* of two single-stranded regions (that is, one strand must run 5′ to 3′, while the partner strand runs 3′ to 5′).
- *Single-stranded loops* can form in the molecule; they must be *three or more* nucleotide bases in length.
- Three or more bases in a single-stranded loop can pair with another single-stranded region elsewhere in the molecule. A three-dimensional (tertiary) shape called a *pseudoknot* results.

These rules are based on observations scientists have made of the folding behavior of biologically active RNA molecules.

Challenge Questions

1. Draw a picture of each RNA's shape.

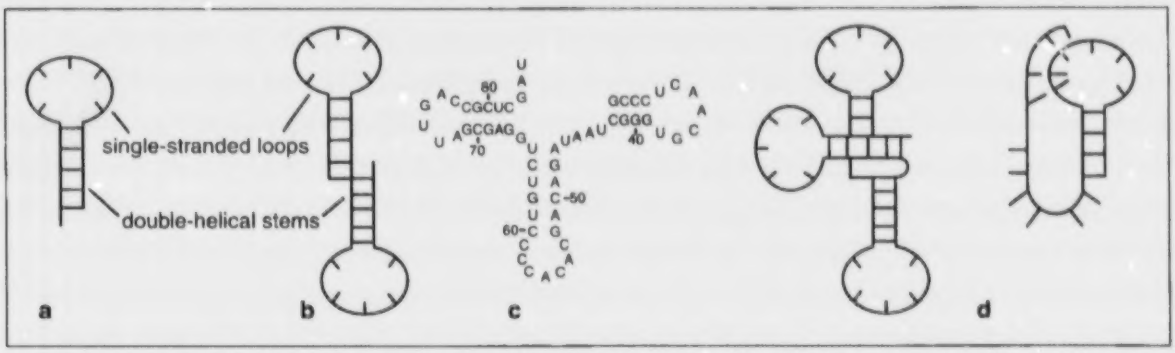


Figure 1.6 Some examples of RNA shapes and structural elements. Single molecules of RNA consisting of one (a), two (b), or three (c) stem-loop elements. (d) A pseudoknot configuration.

2. Write a detailed description of the three-dimensional shape for your RNA sequences, indicating the structural features present. Figure 1.6 depicts some common structural elements of RNAs.

3. Do any of your sequences have more than one possible structure that fits the folding criteria?

FOR YOUR INFORMATION
RNA Structure: Tapes to Shapes

The shapes in Figure 1.6 represent some aspects of RNA structure, but they do not fully capture the potential for three-dimensionality in the molecule. Because RNAs perform many different tasks in the cell, a variety of different and complex shapes is required. Like proteins, complex RNA shapes arise through interactions at several structural levels (Figure 1.7).

The order of the nucleotide monomers in the tapelike *primary* (1°) *structure* of RNA or DNA forms a code, both for storing genetic information and for determining the molecule's shape. In DNA, two linear molecules intertwine into the familiar double-stranded helix, with complementary nucleotide bases of adjacent strands pairing by hydrogen bonds. RNA is more likely to exist as a single molecule; nevertheless, hydrogen base interactions are important here as well.

The primary sequence of RNA is highly flexible, and thus can bend and fold on itself. This folding permits bases within the *same* molecule to approach one another and pair via hydrogen bonds. Base pairing within an RNA molecule obeys pairing rules similar to those that operate between the two strands of DNA: C pairs with G, and A pairs with U (T in the case of DNA). This folded level of structure is called *secondary* (2°) *structure*. The monomers involved in this base pairing may be widely separated in the primary sequence, but folding of the molecule brings them close enough to bond. Many shapes can result, resembling a hairpin, a stem with a large loop, even a clover leaf. Note that base pairing within an RNA molecule produces local double-stranded regions that take the shape of a helix, similar to a DNA helix. Regions not involved in base pairing remain single-stranded.

Scientists now realize that even higher levels of structure occur in RNA. Like proteins, *tertiary* (3°) and *quaternary* (4°) interactions are possible. For example, a single-stranded loop region in an RNA may fold

FOR YOUR INFORMATION (continued)

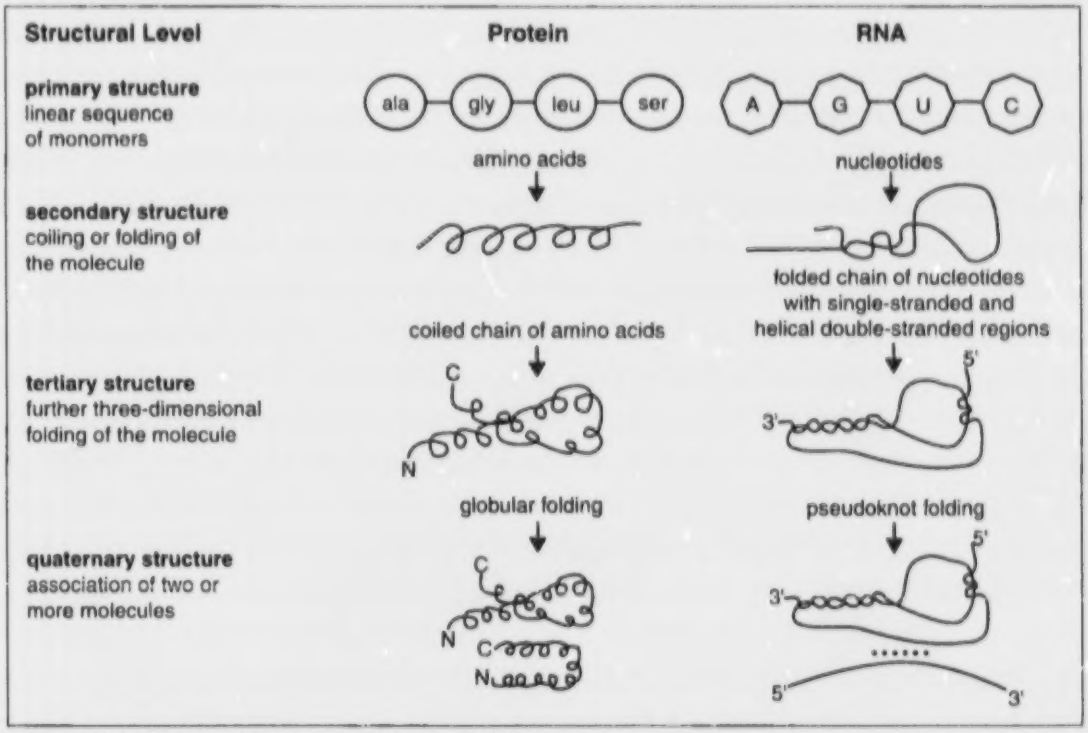


Figure 1.7 Comparison of RNA and protein structural elements

over and base-pair with some other single-stranded region of the molecule. This forms a three-dimensional shape known as a *pseudoknot*, an example of tertiary structure in RNA. It is even possible for two different RNA molecules to interact via base pairing between their single-stranded regions, an example of quaternary structure.

Exercise 1.2: Alternative Modeling Method

In this part of Activity 1, you will use a simple molecular model and a specific set of folding rules to discover and visualize some common RNA shapes. You will fold a simple linear model of the RNA sequences in Figure 1.3 into more complex shapes by applying folding rules.

Materials

- vinyl tubing
- preprinted strips of paper
- twist-ties

Procedure

1. Each team member receives two preprinted strips of paper, each bearing a *different* base sequence listed in Figure 1.3, as follows: Two members *each*

BEST COPY AVAILABLE

139

receive a copy of sequence 1 and a copy of sequence 3; the other two members *each* receive a copy of sequence 2 and a copy of sequence 4. The 3′- and 5′-ends of the molecule are indicated on the strip.

2. Each team member also receives *two* 50-cm lengths of clear vinyl tubing and 10 twist-ties. Insert each of your paper-strip sequences into a separate tube. (Allow the end tab that bears the sequence's number to protrude from the tube.) Hint: Insert the strip slowly while keeping the tube straight. This is easily done by holding the tube vertically while securing one end under your foot.

3. Use the folding rules listed in Figure 1.5 to fold your sequences. Hint: Work in parallel with the team member using the same sequences as you, so that you can help each other. Use twist-ties to represent a few of the hydrogen bonds and to stabilize your model.

4. Compare your structures with those of your teammates and with those of other teams. Reject any invalid structures that do not conform to the folding rules. The valid models represent the potential three-dimensional structures dictated by your RNA sequence.

Challenge Questions

1. Draw a picture of each RNA's shape.

2. Write a detailed description of the three-dimensional shape for your RNA sequences, indicating the structural features present. Figure 1.6 depicts some common structural elements of RNAs.

3. Do any of your sequences have more than one possible structure that fits the folding criteria?

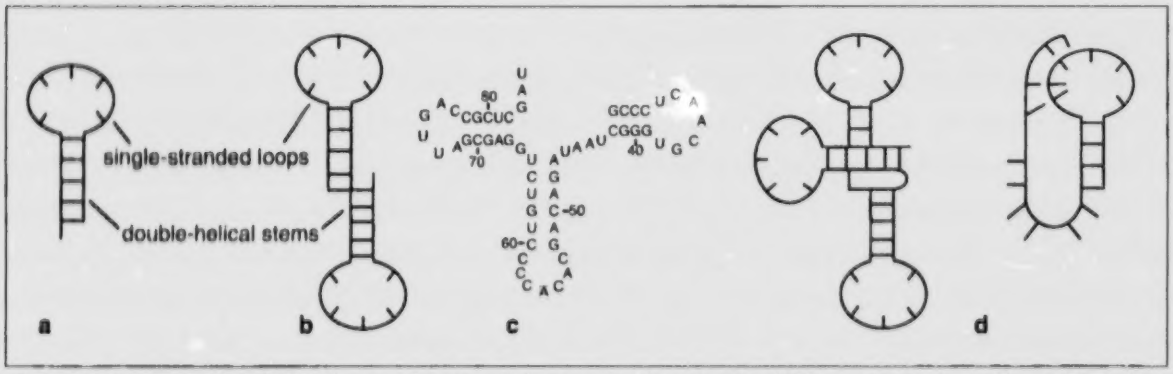


Figure 1.6 Some examples of RNA shapes and structural elements. Single molecules of RNA consisting of one (a), two (b), or three (c) stem-loop elements. (d) A pseudoknot configuration.

140

Exercise 1.3a: Finding Function

Now we will relate RNA structure to function by determining which of your four RNA sequences is the biologically active structure capable of carrying out a specific simulated molecular task, in this case, binding to a particular "protein."

Materials

- models from Exercise 1.2
- styrene ball

Procedure

1. After your team has agreed on valid models for each sequence from Exercise 1.2, arrange one set of models on a desk. Each team receives a 5-cm diameter styrene ball that simulates a small globular protein molecule. Note that the ball has several letters marked on it. Each letter represents the position of a different amino acid in the primary sequence of the protein. When you align the protein model with the RNA model that has the correct "functional" base sequence, the letters on each model will match up identically, indicating a functional fit.

2. Working as a team, test each of your RNA models for its ability to "bind" the "substrate protein."

Challenge Questions

1. For which sequences did you find a function?

2. For sequences having more than one possible shape, did both shapes have the same functional ability? What does this imply about the relationship between molecular structure and function?

3. Can you think any functions performed by RNAs? List them.

Exercise 1.3b: Effects of Mutation

Now you will determine whether a mutation (change) in the primary base sequence of the RNA molecule will alter the shape and function of the molecule.

Materials

- models from Exercise 1.2
- styrene ball from Exercise 1.3a
- dull tape
- colored marking pens

Procedure

1. Your instructor will give you information showing a mutation in your primary RNA sequences. Using your model, determine whether the mutation has any effect on the structure of your RNA molecule. Hint: You can cover the section of the model to be mutated with dull tape and use a colored marking pen to change the base to fit the mutation data.

141

2. Test each of the mutated sequence models for function using the "protein" ball from Exercise 1.3a.

Challenge Questions

1. Write a brief description of any structural effects that the mutations had on your model RNAs, such as "It changes the length of a loop" or "It eliminates a double-stranded stem." Indicate in your description which RNA sequence (1–4) you are describing and the base changes involved.

2. Do the mutations alter the function of the sequence(s) in the "ball test"? How?

3. After reading the FYI essay *Molecular Selection in Real Time*, in Activity 3, discuss with your teammates any general similarities between the approach to RNA structure and function that you employed in Exercises 1.2 and 1.3a and the recently developed laboratory approach called *in vitro* nucleic acid selection.

Exercise 1.4: Using the Computer to Determine RNA Structures

Having now spent some time deciphering the structures of quite short RNA molecules, you can appreciate how complex base pairing is in the more typical molecules isolated from organisms, which are hundreds to thousands of nucleotides in length. Here is where the computer is put to good use, rapidly determining the alternative shapes that a molecule is capable of assuming and calculating which of these shapes is energetically the most likely.

We will use an RNA folding program called *mfold*, which is available on the World Wide Web.

Materials

- access to the World Wide Web

Procedure

Access the mfold program at http://www.ibc.wustl.edu/~zuker/rna/form1.cgi. Apply the program in turn to each of the sequences in Figure 1.3 that you folded. The program is easy to use, and you can simply follow the on-screen instructions to analyze your sequences. Following are some points to keep in mind as you use the program:

1. The program analyzes only one molecule at a time; you must go through the entry and analysis process for each sequence. Once the program has analyzed and displayed a molecule, you can simply delete that sequence from the data entry field and enter a new sequence to be analyzed.

2. Enter sequences beginning with the 5′-most base.

3. Use the default parameters displayed on screen (for example, *the molecule is linear* or *fold at 37 degrees*), with the following exceptions: choose *image resolution low* and *structural format: bases*.

BEST COPY AVAILABLE

142

4. Once you have entered the sequence and analysis parameters, scroll to the bottom of the page and click on *Fold RNA*. Folding and display typically require 1–2 minutes, depending on how busy the server is.

5. When the *Output* page appears, scroll down to the links labeled *sequence*. When more than one shape is possible, a numbered *sequence* link will appear for each. Click on any one of the numbered sequence links to open the page displaying the folded RNA structures.

Challenge Questions

1. Did your predictions about the shapes of the sequences match those of the computer program?

2. Did the program reveal more than one shape for any of the molecules? If so, which shape is more likely on energetic grounds?

Analysis Questions

1. Review your answers to each Challenge Question and discuss the connection between structure and function for RNA.

2. If an RNA molecule functions as a catalyst, how might the folding pattern be important for its activity?

3. In a population of related molecules, could an altered (mutant) structure ever have an advantage? Explain.

4. Do you think molecules (or molecular function) can be subject to natural selection? Explain.

5a. Compare the consequences for a species between two changes potentially affecting its RNA structure across time: (1) a base alteration that might occur randomly in the RNA after it has been synthesized *versus* (2) a random mutational change that can be traced to the template for the RNA (such as the DNA gene that encodes it).

5b. Would your answer in Question 5a change if the RNA molecule itself were capable of self-replication? Explain your response.

143

Activity 2
RNA Catalysis

Introduction

When someone asks you, "What does RNA do?" you probably think of messenger RNA (mRNA), which is copied from a gene and communicates the information needed to make a protein. In this familiar role, RNA is doing a job similar to DNA, storing genetic information. This information-storage function is analogous to the way a compact disc (CD) stores musical information. To use or copy the information in a CD, however, you need a form of technology capable of performing these functions: a CD player. However, in terms of this analogy, RNA is a surprising molecule. It's a bit like a music CD that needs no CD player and can play itself. In other words, RNA has dual functions: Not only can it *store* genetic information, but in some cases it can also *transfer* and *use* this information by acting as a biochemical *catalyst*.

A catalyst is a molecule that increases the rate of a chemical reaction. Catalysts are essential to living systems because they allow biochemical reactions to occur fast enough to be compatible with life. Biochemical catalysts speed up reactions by physically binding the reacting molecules, bringing them near one another, and orienting them in just the right way to make the reaction more likely to proceed. Chemists describe this "matchmaker" function of the catalyst as a lowering of the *activation energy* of the reaction, like lowering an energy "hill" that reactants must get over before they can continue along the reaction path. The same reaction could happen without the catalyst, but it might take years instead of milliseconds to complete. Biochemical catalysts also make reactions more specific by bringing together particular molecular reactants.

The most familiar biological catalysts are enzymes. Enzyme catalysts are made of protein, and today they catalyze the vast majority of biochemical reactions, from synthetic reactions like making nucleic acids and other proteins to degradative reactions like releasing energy from sugars and fats.

Some RNA molecules also are catalysts; they are called *ribozymes*. As catalysts, ribozymes appear to function by physical mechanisms similar to those of protein-based enzymes. However, the types of reactions catalyzed by ribozymes are limited. In nature, known ribozymes primarily catalyze reactions on other nucleic acids and, very likely, some key reactions involved in protein synthesis. No doubt other ribozyme-catalyzed reactions are yet to be discovered. In the laboratory, scientists recently synthesized several novel RNA and single-stranded DNA molecules not known to occur in nature that also act as catalysts.

BEST COPY AVAILABLE

It may surprise you to learn that RNA (and even single-stranded DNA) can be a catalyst; this fact certainly surprised scientists when it was first discovered. At first glance, RNA appears to have little in common with proteins to account for its catalytic ability. On closer inspection, however, some critical similarities appear: Both proteins and RNA have subunits containing chemically reactive groups of atoms (for example, the imidazole group found in certain amino acids and nucleotides), and both have complex shapes. A key point: Closely related objects (like RNA and DNA), and even not-so-closely related objects (like RNA and protein), that share key features of structure and shape can frequently perform similar functions.

In living cells, RNA molecules carry out a variety of important tasks, some catalytic and some not. For example, RNA participates in the replication of DNA, functions as the genome of many viruses, and participates in the cellular localization of newly made proteins. Another important role is in the maturation of mRNAs. mRNAs are copies of the genetic information encoded in DNA. But before an mRNA copy can be used to direct protein synthesis, it must undergo a biochemical maturation process that removes noncoding base sequences, known as *introns*. In eukaryotic cells, intron removal is accomplished by several small RNAs plus some proteins that together make up an mRNA "splicing machine"called a *spliceosome*. A few RNA introns, however, are so versatile that they can actually splice themselves out of the precursor RNA with no help from proteins. The surprising discovery of these self-splicing RNA introns by researcher Thomas Cech in 1982 first revealed RNA's catalytic ability and earned Cech a Nobel Prize.

Once a mature mRNA is formed, the subsequent steps in the synthesis of proteins rely heavily on several additional RNA molecules, from the *transfer RNAs* (tRNAs) that carry the amino acids, to the *ribosomal RNAs* (rRNAs) that, together with a variety of protein molecules, make up the ribosome itself. You recall that the ribosome is the cellular structure where amino acid monomers are linked together using the directions encoded in the mRNA to form proteins. About 80 percent of the RNA in a cell is ribosomal, and there is recent evidence to suggest that one of these ribosomal RNAs (the 23S rRNA) is in fact a ribozyme catalyst that chemically couples the amino acids during protein synthesis.

Several classes of naturally occurring ribozyme RNA have been identified in nature, and hundreds of individual examples have been found in a wide variety of living cells (Figure 2.1). Still other ribozymes have been constructed in the laboratory by conducting *molecular evolution in a test tube*. In this process, scientists start with a large, random population containing many different laboratory-synthesized RNA sequences. The researchers select from this population only those RNAs that have particular, desired functions, such as the ability to bind a particular target molecule or catalyze a particular reaction. These synthetic molecules reveal many previously unsuspected capabilities of nucleic acids, including the ability of RNA to make partial copies of itself. We will explore the important topics of *in vitro* molecular evolution and RNA self-replication in Activity 3.

You can find out more about RNA catalysis and other newly recognized cellular roles of RNA in the references listed in Additional Information. Particularly useful is a somewhat advanced book, *The RNA World*. A set of videotaped lectures about RNA by Nobel Prize winner Thomas Cech is also very informative.

Figure 2.1 Naturally occurring ribozymes.

Class	Size	Reaction	Source
group I intron	large: 413 NT in *Tetrahymena thermophilia*	intron excision	eukaryotes, eubacteria, and viruses
group II intron	large: 887 NT in yeast mitochondria	intron excision	eukaryotic organelles and eubacteria
RNase P	large: 350–410 NT	hydrolytic endoribonuclease	RNA subunit of eubacterial RNase P
hammerhead	small: 31–42 NT (enzyme strand can be 16 NT)	RNA cleavage	viral satellite RNA in plants, viroids, and newt satellite DNA
hairpin	small: 50 NT (minimum sequence)	RNA cleavage	(-) strand satellite RNA of tobacco ringspot virus
hepatitis Delta virus (HDV)	84 NT (required)	RNA cleavage	HDV
Neurospora VS RNA	881 NT (164 sufficient)	RNA cleavage	*Neurospora* mitochondria

Additional Information

Cech, T.R. (1998). Structure and mechanism of the large catalytic RNAs: Group I and group II introns and ribonuclease P. In Gesteland, R.F., & Atkins, J.F. (Eds.), *The RNA world* (pp. 239–269). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Landweber, L.F., Simon, P.J., & Wagner, T.A. (1998). Ribozyme engineering and early evolution. *BioScience, 48*(2): 94–103.

Howard Hughes Medical Institute Holiday Lecture on Science. (1995). *The Double Life of RNA* [videotape set]. A four-tape series of lectures by Nobelist Thomas Cech, discussing his discovery of ribozymes and other important functions of RNA. Summaries of the lectures can be found online at http://www.hhmi.org/.

Challenge Questions

Read the Introduction, then discuss any general similarities between catalysis by proteins and nucleic acids.

Exercise 2.1a: Modeling the Hammerhead Reaction

In this activity, you will use pipe cleaner models to model the catalytic activity of two RNA molecules: a simple catalytic RNA called the hammerhead ribozyme and the first catalytic RNA to be discovered, the self-splicing group I intron.

First you will examine the catalytic activity of the so-called hammerhead ribozyme. The name derives from the superficial resemblance of this RNA's secondary structure, as usually depicted, to a carpenter's hammer. Hammerhead RNA sequences are embedded in the RNAs of many viruslike agents that infect plants. The RNA

BEST COPY AVAILABLE

genome of these so-called viroids and virusoids is replicated in the form of a long precursor transcript made up of several tandemly repeated genomes plus repeats of the smaller hammerhead sequence. The hammerhead sequences catalytically cut themselves out of the long precursor transcript, in the process cleaving it into many individual genome-length segments. The hammerhead RNA sequence is thus an *endoribonuclease*, a catalyst able to cleave RNAs internally.

Materials

- pipe cleaner model of hammerhead ribozyme
- pipe cleaner model of substrate RNA
- scissors
- tape

Procedure

The pipe cleaner model with four clusters of labeled bases represents a hammerhead RNA sequence that has been genetically engineered to interact with an external RNA substrate. (An external substrate is an RNA molecule other than the one in which the hammerhead resides.) The other labeled pipe cleaner represents the substrate sequence for this hammerhead ribozyme.

1. Use the base sequences in Figure 2.2 to locate and mark the 5′-end of each pipe cleaner model (use a black ink dot or marking of your choice).
2. Use the sequences on your hammerhead and substrate pipe cleaner models to fold them into the appropriate secondary structures.
3. Identify a base-pairing interaction *between* the hammerhead and substrate molecules.
4. Locate the enzymatic cut site in the substrate. This site has the general sequence NUH, where N can be any nucleotide, and H can be A, C, or U.
5. Cut the substrate on the 3′ side of the base you identified as H.

Challenge Questions

1. Describe what this reaction accomplished. Draw the resulting molecules.
2. Suggest a practical application for such an endonuclease.

As you can see, formation of the hammerhead ribozyme's catalytic site depends on the molecule first folding into the correct structure.

hammerhead ribozyme
3′_____UUCUUCA_____CAGG_____CCUG_____UUGCA_____5′

substrate
5′_____GGCAAGAAGUCAACGUGUGGG_____3′

yellow = A green = U black = G white = C

Figure 2.2 Base sequences.

Recall that the naturally occurring hammerhead sequences in the plant viroids and virusoids typically cut only the RNA molecule in which they are embedded. In contrast, hammerhead sequences like the one you just modeled, which can cleave *external* substrates, have for the most part been fashioned in the laboratory through genetic engineering. These engineered ribozymes function like protein enzymes because they react with many copies of an external substrate in a multiple turnover process. To develop a potential form of gene therapy, researchers are attempting to design RNA-cutting RNAs like these, which can recognize and cut either foreign RNAs (such as the HIV viral genome) or mutated cellular RNAs (such as the altered RNAs that can promote cancer). The hope is that when these ribozymes are incorporated into affected cells, they will specifically eliminate the offending target RNAs.

Exercise 2.1b: Modeling the Group I Intron Reaction

Next, you will model the first catalytic RNA discovered, the self-splicing group I intron. This RNA was first found in 1982 by Thomas Cech in the nuclear ribosomal RNA of the protozoon *Tetrahymena thermophilia*. Introns of this type also occur in other eukaryotes as well as in prokaryotes and viruses. The group I intron RNA is said to be *autocatalytic* because it acts on itself. The intron resides within a larger precursor rRNA molecule and catalyzes a two-step reaction that dramatically alters its relationship to the larger molecule. Like other functional RNAs, group I introns must first fold into a precise secondary structure by internal base pairing. Proper folding enables the intron to recognize and react with particular bases at the *splice sites* within the larger RNA.

Materials

- pipe cleaner model of RNA
- paper model of guanosine molecule
- paper model of an OH group
- tape

Procedure

The pipe cleaner model provided represents a precursor rRNA containing a group I intron sequence flanked on either side by exon sequences. The 5'- and 3'-ends of the molecule are indicated, and several of its nucleotides are color-coded as follows:

yellow = A green = U black = G white = C

The following key bases are marked in the model:

5'___CCCUCUA_____UUUA_____AGAGGG_____GU_____3'

1. Locate the following landmarks in the model:

   - The 5'-most exon, which *ends* with the base sequence CCCUCU.
   - The 3'-most exon, which *begins* with and includes the U of a marked G-U pair.
   - The self-splicing intron is everything in between these exons.

2. Use the sequence on the model to form an appropriate secondary structure.

BEST COPY AVAILABLE

3. Simulate a nucleophilic attack and cleavage by the 3′-OH group of the free guanosine. This cleavage occurs at the 5′ exon-intron boundary of the RNA (at a PO_4^- group not shown in the model). Simulate cleavage by cutting the pipe cleaner *between* the U and the A at the 5′ exon-intron boundary.

4. After cutting, attach a paper OH group to the 3′-most base of the 5′ exon; this creates a reactive OH group at the end of the 5′ exon. Next, attach the paper guanosine by its 3′-OH to the 5′-most end of the intron.

5. Simulate the second step of the self-splicing reaction by using the 5′ exon's exposed OH to attack and cleave at the 3′ exon-intron boundary. Simulate this by making a cut between the G-U pair at this boundary. After cutting, transfer the 5′ exon's exposed OH group to the 3′-most end of the intron.

6. Bring the 3′-end of the left exon next to the 5′-end of the right exon and tape them together.

Challenge Questions

1. Describe what these reactions accomplished. Draw the resulting molecules.

2. What is the base sequence at the junction of the spliced exons?

3. Where did the guanosine end up?

4. Which catalytic ability does the group I intron possess that is lacking in the hammerhead? in the ribozyme?

Exercise 2.1c: The Group I Intron RNA Can Go a Step Further

After removing itself intact from the larger rRNA precursor, as you have just demonstrated, the group I intron often continues its cleavage-ligation activity by reacting on itself internally. In this secondary, follow-up reaction, a reactive nucleophile group at the 3′-end of the intron (the chemical nature of this group should be familiar by now) cleaves the intron at a particular purine-pyrimidine junction.

Use the reactive 3′-OH of the intron as a nucleophile to carry out another cleavage-ligation reaction. Remember, the nucleotides are color-coded as follows:

yellow = A green = U black = G white = C

1. Locate the only remaining purine-pyrimidine junction marked in the intron model.

2. Simulate the intron's cleavage of itself by cutting it between the purine and pyrimidine bases.

3. Transfer the intron's 3′-OH to the pyrimidine base newly exposed at the end of the shorter fragment.

4. Simulate ligation within the intron by taping its free ends together.

Challenge Questions

1. What is the end result of this reaction? Draw the resulting molecules.

2. Where does the original guanosine molecule end up?

As you have seen, the group I intron is self-splicing, catalytically removing itself from within the larger precursor rRNA in which it resides. This self-splicing occurs by a two-step reaction sequence: a *cleavage* reaction, which excises the intron from between the two coding exons that flank it, followed by a *ligation* reaction, which splices the exons together to create the mature, and now smaller, rRNA. These cleavage-ligation reactions require no source of chemical energy, such as ATP or GTP, because they simply entail the replacement of one phosphodiester bond with an energetically equivalent one (such reactions are termed *transesterifications*). All that is required *in vitro* for self-splicing by the group I intron is free guanosine (or any of its phosphorylated derivatives: GMP, GDP, or GTP) plus a divalent cation (typically Mg^{++}). Group II introns, found in cyanobacteria and eukaryotic organelles such as mitochondria and chloroplasts, have simpler requirements for their self-splicing and do not need a free nucleotide.

Naturally occurring self-splicing introns react only with themselves and thus not with external substrates. This self-limited reactivity stands in contrast to the repeating action of protein enzymes, which typically react with many copies of an external substrate in a repetitive turnover process. Some ribozymes of the nonintron type (such as ribonuclease P, which participates in the cleavage and maturation of transfer RNAs), are in fact capable of reacting with multiple copies of an external substrate.

Exercise 2.2: A Laboratory Study of the Self-Splicing RNA from *Anabaena*

In this laboratory experiment, you will carry out in the test tube two important reactions involving RNA. The first is transcription, a classic reaction in which a DNA sequence is copied into a complementary RNA sequence. The RNA sequence that you will transcribe contains a catalytic group I intron like the one you modeled in the previous exercise. In the second part of this experiment, you will examine the autocatalytic self-splicing activity of this RNA. The discovery of catalytic RNAs, or ribozymes, of which the group I intron is just one example, resulted in a joint Nobel Prize for Thomas Cech and Sidney Altman in 1989.

In most eukaryotic genes and some prokaryotic and phage genes, the genetic information is not colinear with the chromosomal DNA sequence. Rather, the information is interrupted by noncoding sequences, some many hundreds of bases long. Chromosomal DNA is therefore a mix of coding regions, called *exons*, and noncoding regions, called *introns*. During transcription, the exon and intron sequences are transcribed together into a primary RNA transcript. In order to be translated into protein form, this primary transcript needs to be processed into a mature RNA. Processing involves catalytic removal of the introns and the simultaneous splicing together of the exons. Such *post-transcriptional processing* is not confined to mRNA production, but is also found in the production of functional rRNA and tRNA.

Four biochemical classes of spliceable intron have been identified. Some of these need accessory proteins and/or an energy source to catalyze splicing efficiently, but

BEST COPY AVAILABLE

the so-called group I and group II introns require neither: They are self-splicing. All that these RNAs require for self-splicing is a guanine nucleotide cofactor (guanosine, GMP, GDP, or GTP) and a divalent cation (Mg^{2+}). The RNA you will use is a pre-tRNA from the cyanobacterium *Anabaena* and contains a group I intron.

You will transcribe this autocatalytic RNA *in vitro* from an engineered plasmid (pAtRNA-1) that contains a cloned copy of the $tRNA^{Ile}$ DNA gene sequence (PCC7120) from *Anabaena*. This tRNA-encoding DNA was cloned into a commercial plasmid (pBS-) downstream from a copy of the T7 promotor. The promotor provides a transcription start site for the enzyme T7 RNA polymerase, which you will use to copy the insert DNA into a functional RNA *in vitro*. The Cech lab further engineered the insert DNA at its 3'-end to include an *EarI* site, a particular restriction endonuclease site. Cutting open the circular plasmid with the *EarI* enzyme at this site linearizes the plasmid and provides a transcription termination site for the T7 RNA polymerase. This type of termination is called *run-off transcription*, as the polymerase literally falls off the end of the linear DNA when it reaches the *EarI* site. The linearized plasmid will yield a primary RNA transcript of 334 nucleotides in our experiment.

The self-splicing reactions that form the mature $tRNA^{Ile}$ involve both cleavage and ligation. A schematic of the reaction is shown in Figure 2.3. Note that it is a two-step

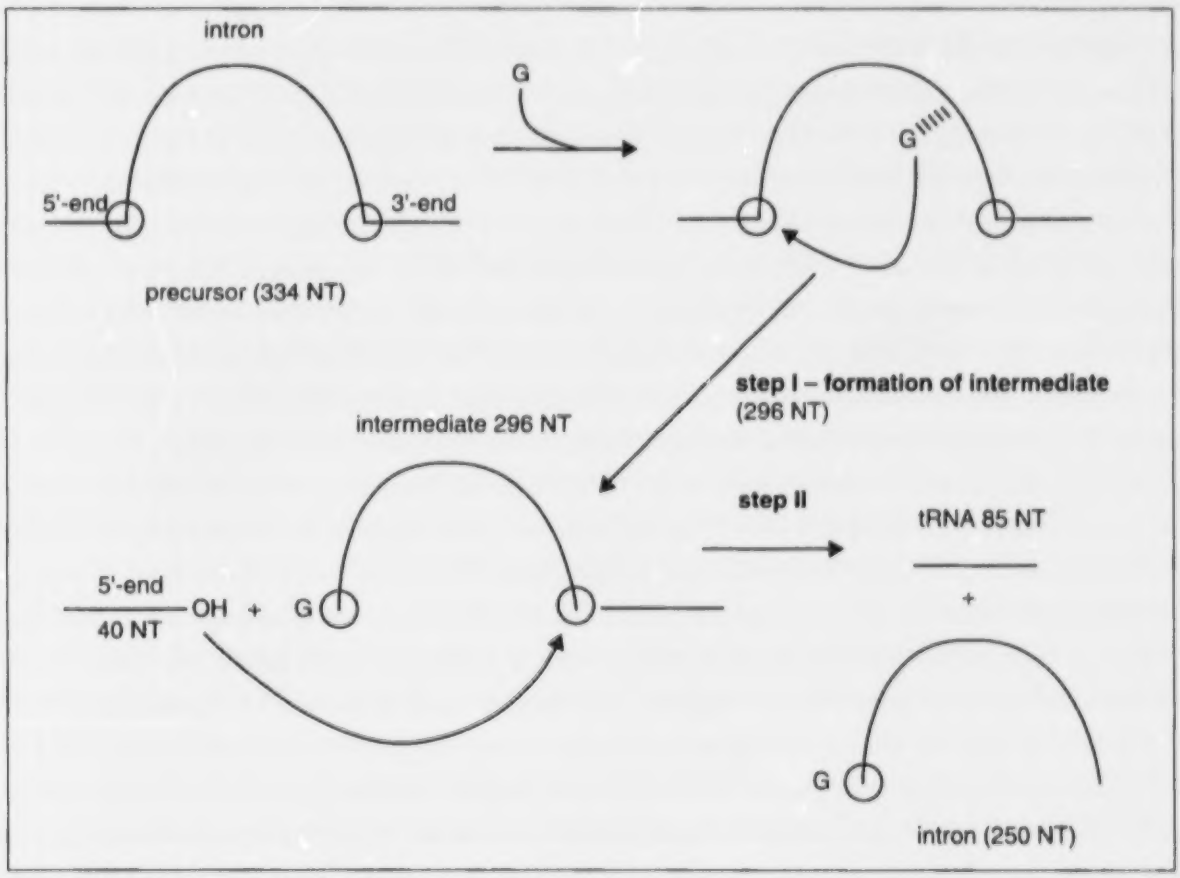


Figure 2.3 Self-splicing reaction of the *Anabaena* group I intron.

mechanism in which different-sized fragments are produced during each step. Initially, we have the primary transcript with a length of 334 bases. In the first reaction step, exogenous guanosine functions as a nucleophile to attack and cleave the intron at its 5′-end. The resulting cleavage of the RNA at this 5′ exon-intron boundary yields a 296 NT intermediate (plus a small 40 NT fragment representing the 5′-end of the exon). The second step, attack of the 5′-end of the exon on the 3′ intron-exon boundary, cleaves the RNA at this boundary, ligating the 5′- and 3′-ends of the exon into a mature 85 NT tRNA and releasing the free 250 NT intron. The overall result of the cleavage-ligation reactions is the release of the 250 NT intron plus the mature 85 NT tRNA molecule.

These reactions are simple transesterifications (involving the replacement of one phosphodiester bond with an energetically equivalent one) and thus require no source of chemical energy. We can study the kinetics (time-dependent changes) of such a two-step reaction and determine which of the steps is slower (rate limiting). This can be done by simply observing the pattern of splicing products. If the first step is faster, then the intermediates made in the first step (296 NT and 40 NT) will accumulate in the reaction mix. If the second step is faster, then you will see few intermediates, only the 250 NT and 85 NT final products.

You will use acrylamide gel electrophoresis to analyze the kinetics of this RNA splicing system. By running aliquots of the reaction mixture at various times during its progress, you should be able to inspect the gel and determine the relative concentrations of the different fragments produced during the reaction. This qualitative inspection allows one to determine which step is faster.

Background and Experimental Procedure

DAY 1, TRANSCRIPTION REACTION AND RNA ISOLATION

In the first part of the experiment, you will transcribe *in vitro* a linear version of the pAtRNA-1 plasmid. This recombinant plasmid encodes the catalytic intron sequence of interest and was linearized by cutting the plasmid with the *EarI* restriction enzyme. You will use this plasmid DNA as the template in the transcription reaction. To prevent the intron from self-splicing prematurely as it is being transcribed (rather than in the second part of this experiment), you will carry out the reaction under conditions of high nucleoside triphosphate (NTP) concentration. The reason is that self-splicing requires binding of a guanosine molecule to the folded precursor RNA (analogous to a substrate binding to an enzyme). The binding of guanosine, in turn, is promoted by Mg^{+2}, and so the negatively charged NTPs are present in the reaction mix to sequester this positively charged magnesium, thereby limiting the guanosine's ability to bind to the intron and stimulate its catalytic action.

After the transcription reaction, you will treat the mix with DNase (RNase-free) to remove all the high-molecular-weight plasmid DNA. You will then isolate the primary RNA transcript from any contaminating DNA fragments by binding the RNA to a silica-gel matrix, from which it will be eluted in purified form with water. You will perform both transcription and RNA isolation on the first day of the experiment.

DAY 2, SELF-SPLICING REACTION AND GEL ELECTROPHORESIS

In this part of the experiment, you will add Mg^{+2} to the purified RNA solution to allow it to self-splice. As self-splicing proceeds in the test tube, you will remove

aliquots of the reaction at various times and quench them. You will then determine the size of the RNA fragments present at each time point by assaying them electrophoretically on an acrylamide gel. The gel separates molecules according to size, with smaller molecules moving farther. By visualizing the RNA fragments on the gel and comparing their migration distances with those of a set of standard RNAs of known size, you can assign sizes to the RNA fragments. You thus have an electrophoretic assay for the splicing reaction. You will relate the observed electrophoretic pattern of RNA fragments to one of two possible mechanisms for the two-step self-splicing reaction: first step faster or second step faster.

Procedure

DAY 1, TRANSCRIPTION REACTION AND RNA ISOLATION

Your instructor will provide you with *EarI*-cut plasmid DNA (approximately 2 µg/µl concentration) as well as the reagents necessary for *in vitro* transcription. Use sterile technique in this experiment.

CAUTION: Wear gloves at all times.

Transcription Reaction

1. Add the indicated number of microliters* of the following reagents to a sterile Eppendorff tube.

DNA	Buffer**	H_2O	rNTP Mix***	RNase Inhibitor	T7 RNA Polymerase
20	5	48	10	2	5

* These volumes assume that you are not using a commercial transcription kit.
** 1 M Tris, pH 8.0
*** Prepare ahead by mixing equal volumes of the four 100 mM rNTP stock solutions. This mix provides a 4.0 mM final concentration of each rNTP in the reaction tube (16 mM in total rNTP).

2. Mix the reagents with a pipette tip and incubate the tube at 40°C for 2 hours.

3. Following the first incubation, add 1 µl of RNase-free DNase to the tube. Mix and incubate for an additional 30 minutes at 40°C.

Isolation of RNA

After completing the transcription reaction, you will isolate the transcribed RNA for use later in the self-splicing reaction. You will use a commercial RNA isolation kit that employs a silica-gel column and provides all the necessary reagents.

1. Add 50 µl of sterile water to the transcription mix tube.

2. Add 350 µl of RLT solution and 250 µl of 95% ethanol to the tube and mix by pipetting up and down.

3. Load the entire volume on the RNA isolation column provided.

4. Microfuge for 15 seconds.

5. Wash the column with 500 µl of RPE and collect the filtrate in a new Eppendorff tube. Watch the volume of wash solution accumulating in the collecting

tube and dump it out when the fluid level in the RNA column gets close to the bottom of the column. You need to do this periodically, as the combined volume of the washes exceeds the volume of the collecting tube. Wash once more with 500 µl RPE. When the column has drained, microfuge it for 2 minutes to remove residual wash solution.

6. Get a sterile Eppendorff and elute the bound RNA into it with 50 µl of DEPC-treated water. Store the RNA solution in the -70°C freezer until it is needed for the self-splicing reaction.

DAY 2, SELF-SPLICING REACTION AND GEL ELECTROPHORESIS

Now you will analyze the time course of the self-splicing reaction. You will initiate the splicing reaction by adding guanosine and a buffer containing Mg^{+2} to the purified catalytic RNA. As the reaction progresses, you will remove small aliquots at various times and dilute them into solution that stops the reaction of that sample. You then will run samples stopped at various times during the reaction on a polyacrylamide gel containing urea. Urea is a denaturant that eliminates RNA's three-dimensional structure. With no secondary or tertiary structure to complicate matters, the RNA fragments separate simply according to their molecular size. Based on the expected sizes of each fragment, you should see the following bands in an ideal separation:

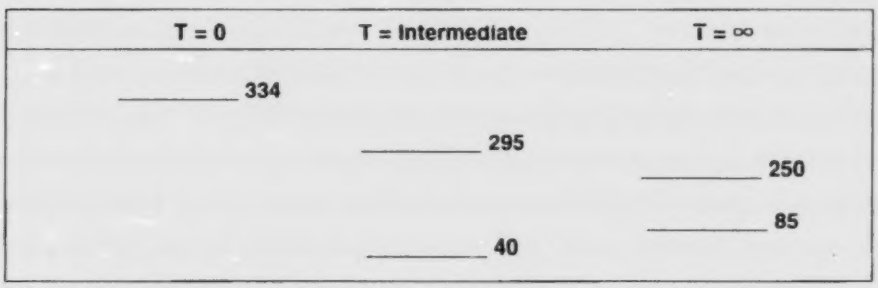


Figure 2.4 Distribution of bands expected on the acrylamide gel.

Splicing Reaction

1. Set up four Eppendorff tubes, each containing 10 µl of stop solution. Label the tubes as follows: T = 0, T = 30, T = 60, and T = 90.

2. Set up a reaction tube containing 44.5 µl of the isolated RNA and 5 µl 10X HEPES buffer (with Mg^{++}). Mix and incubate at 50°C for 10 minutes. (Can you think of a reason for this preincubation?)

3. Begin timing as soon as you transfer the reaction tube to a water bath at 32°C. At the 1.5 minute mark, remove a 10 µl aliquot from the reaction tube and add it to the T = 0 stop solution tube.

4. At the 2.0 minute mark, add 5 µl 10X guanosine to the reaction tube and mix rapidly at 32°C. Every 30 seconds, transfer a 10 µl aliquot of the reaction to the appropriately labeled stop solution tube.

Run the time-point samples on the polyacrylamide gel as described in the following section.

BEST COPY AVAILABLE

Polyacrylamide Gel Electrophoresis

Your instructor will demonstrate how to assemble the gel apparatus that you will use. You may use a prepoured gel, or you may pour your own gel. To pour a 9-by-11-inch gel, you will need to prepare about 120 ml of acrylamide solution.

1. Dissolve 7.2 gm of acrylamide powder in 120 ml of 1X TBE buffer containing 8 M urea.

CAUTION: Wear a mask and gloves when handling acrylamide powder.

2. Add 1% by volume of freshly prepared 10% ammonium persulfate and 0.1% by volume of TEMED to the acrylamide-urea solution, mix, and carefully pour the solution between the glass plates. Insert the comb. Let the gel polymerize for at least 2 hours.

3. After polymerization, fix the gel plates into the electrophoresis unit, cover with an electrode buffer, and remove the comb. Immediately wash out the wells with a syringe containing electrode buffer. This creates good wells.

   Prerun the gel with no samples for 30 minutes at 30 W, washing out the wells again after the prerun.

   Load the samples. Electrophorese at 30 W, 1,500 V until the slowest moving band, the xylene cyanol, is about 1 to 2 inches from the bottom of the gel.

4. Turn off the power supply, then remove the gel and stain for 30 minutes in ethidium bromide. Photograph the gel.

CAUTION: Observe appropriate precautions around electrophoresis equipment and when handling ethidium bromide.

Challenge Questions

1. Did transcription occur? How do you know?

2. Did splicing occur? How do you know?

3. Explain the gel data in your own words by drawing a diagram of the gel and describing the nature of the bands observed, such as "This band (indicate band with arrow) represents . . . "

4. Which step of the splicing reaction is faster?

Additional Information

Zaug, A.J., McEvoy, M.M., & Cech, T.R. (1993). Self-splicing group I intron from *Anabaena* pre-tRNA: Requirements for base-pairing of the exons in the anticodon stem. *Biochemistry, 32*: 7946–7953.

REAGENT LIST

a. ***EarI*-cut pAtRNA-1** - concentration approximately 2 μg/μl

b. **Sterile 10X transcription buffer** - provided in the transcription kit or prepared as 1 M TrisHCl, pH 8.0

c. **Sterile 0.1 M $MgCl_2$** - autoclaved

d. **Sterile 10X nucleoside triphosphate solutions** - filter-sterilized aqueous solutions containing 25 mM each of rCTP, rUTP, rATP and rGTP

e. **Sterile 5 M NaCl**

f. **Sterile 10X HEPES buffer** - autoclaved 250 mM HEPES buffer, pH 7.5, containing 100 mM $MgCl_2$

g. **Sterile 10X guanosine** - 250 μM guanosine in 1X HEPES buffer

h. **1X Tris-borate-EDTA buffer (TBE)** - 0.1 M Tris base containing 0.083 M boric acid and 1 mM EDTA

i. **Stop solution** - 0.1X TBE containing 30 mM EDTA, 10 M urea, 0.01% bromophenol blue, and 0.025% xylene cyanol

j. **Polyacrylamide gel** - gel made of 6% polyacrylamide plus 8 M urea in 1X TBE

k. **DNase I (RNase-Free)**

l. **T7 RNA polymerase**

m. **RNase inhibitor**

Commercial Kits and Reagents

n. **Plasmid DNA isolation** - QIAGEN Plasmid Mega Kit

o. **RNA isolation kit** - QIAGEN Rneasy Mini Kit

p. **RNA size standards** - (500 to 100 nucleotides) Ambion Century™ Marker Templates

q. ***EarI* restriction enzyme** - New England Biolabs (10 U/μl), includes 10X buffer.

FOR YOUR INFORMATION
The Origin of Life: New Answers for an Ancient Question

The origin of life has always been a fascinating and vexing question. Throughout the history of humankind, this mystery has been the subject of passionate debate, and every age has come up with answers that reflect the religious, philosophical, and scientific beliefs of the time. During the last 400 years, science and technology have been extraordinarily successful in allowing us to understand and manipulate the physical world. More recently, the biological sciences have provided an understanding of and ability to manipulate living systems that is unprecedented in history. In light of such modern advances, it is only natural that we ask the ancient question anew: Where did we, and all life, come from?

For some, the answer is simple, satisfying, and ancient: God or a deity made all life as we know it.

Although the existence of a deity is not a topic that can be addressed through scientific inquiry, modern explanations of life's origin need not be based on received faith or unfounded opinions. We can seek a valid naturalistic explanation for the mechanism of life's origin, an explanation that is consistent with experimental evidence. New data from ongoing scientific research is providing the means to construct informed and testable hypotheses of our origins.

Of the recent data bearing on the origin of life, perhaps the most revolutionary was the discovery in 1981 that RNA can be catalytic. This discovery was significant to the question of life's origin because it showed for the first time that a molecule, RNA, exists that has both of the key properties essential to the beginnings of a living, evolving system: the ability to *encode information* and the *catalytic ability* to conceivably replicate itself and other molecules.

The discovery of catalytic RNA resolved a chicken-egg paradox about whether nucleic acids or functional protein catalysts evolved first, which had long stymied progress in finding a viable scientific explanation for the origin of life. If the nucleic acid RNA can function as a catalyst, then it could in principle have functioned early in evolution as both *genome* (where genetic information is stored as the linear sequence of nucleotide bases) and *replicase* (the catalyst that replicates the genome). No DNA or proteins would be required for an RNA molecule to duplicate itself in this hypothetical RNA world. This RNA molecule could then evolve through natural selection into even more complex RNAs possessing additional coding and catalytic functions.

Eventually in the course of evolution, DNA took over as the primary genetic material, while proteins took over the vast majority of catalytic functions. All organisms today (except the RNA viruses) have a genome of DNA. One reason that DNA may have taken over from—that is, out-competed—RNA is that it is more chemically stable than RNA. RNA slowly breaks down into monomer units even at room temperature, whereas DNA can persist for very long periods, long enough to enable biologists to recover DNA from a 40,000-year-old wooly mammoth (but not from a 70-million-year-old velociraptor). Proteins later emerged as the primary biological catalysts because of the chemical versatility of their 20 amino acid building blocks and their ability to assume a seemingly endless number of potential shapes.

The progenitor RNA almost certainly had a crucial ability conferred by its coding and catalytic properties: the ability to reproduce itself. This property of molecular replication gets to the very heart of what it means to be "living." The distinguished geneticist and evolutionary biologist H.J. Muller, responding to the question of what it means to be a living system, has said, "I think the most fundamental

FOR YOUR INFORMATION (continued)

property distinguishing a living thing . . . is its ability to form copies of itself. We call this 'reproduction' . . . " To further clarify where the origin of life resides, Muller went on to say, "I should draw the line where the Darwinian process of natural selection begins to come in, and that is at the appearance of (molecular) replication of a self-copying kind—that is, the replication of mutations."

Our working hypothesis for the origin of life, the RNA world hypothesis, is that the first "living" and evolving entity is likely to have been a self-replicating RNA molecule. In the following activity, you will have a chance to explore RNA replication further.

Admittedly, naked replicating molecules are a long way from the complexity of even the simplest known cell. But laboratory approaches like those guided by the RNA world hypothesis provide for the first time what has been lacking in origin of life research: an experimentally testable starting point for the first steps in the evolutionary process.

Activity 3
RNA and Evolution

Introduction

CRICK'S CENTRAL DOGMA

In 1968, Francis Crick neatly summarized what was then our understanding of the flow of genetic information in living systems. He proposed what is still referred to as the central dogma of molecular biology. The dogma asserts that in the biological world, the translation of genetic information is fundamentally a one-way process: *from* nucleic acids, where information is stored as the "language" of specific sequences of nucleotide bases, *into* the "language" of proteins, whose functional amino acids and diverse shapes allow them to express (do something with) the information in the cell. Crick's view of the central dogma is summarized in Figure 3.1.

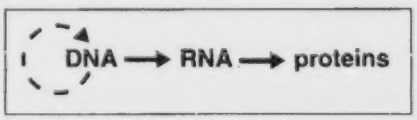


Figure 3.1 Crick's central dogma. The curved arrow indicates DNA's well-known ability to be copied during replication.

The choice of the term *dogma* to describe what is essentially a scientific hypothesis was unfortunate. A dogma, because it implies an unchanging and rigidly held belief, is just the opposite of a scientific hypothesis. A hypothesis is a working proposal that is used to guide future experiments, is held tentatively, and is subject to modification or rejection at any time as new evidence dictates. True to its real nature as a scientific hypothesis, Crick's original proposal has been modified significantly as our understanding of genetic information has grown during the last 30 years.

Crick's main assertion, that genetic information flows from nucleic acids to proteins, and not in the reverse direction, has stood the test of time. No instance of reverse translation from proteins to nucleic acids has ever been observed. But today we know that Crick's original scheme of information flow was incomplete. Newer research reveals that genetic information can flow back and forth *between* the nucleic acids DNA and RNA. The well-known path of transcription of DNA into RNAs has been widened into a two-way street as examples of transcription of RNAs into DNA have been discovered. Some examples of this reverse transcription are the copying of viral genomic RNA into DNA, with the DNA's subsequent insertion into the host cell genome (for example, HIV and other *retro*viruses). Another example is the occasional copying of cellular mRNAs back into DNA form, with their insertion back into the genome (as pseudogenes). The 1970 discovery of the enzyme reverse transcriptase, and the two-way informational "crosstalk" between DNA and RNA that it makes possible, was so significant that it earned David Baltimore and Howard Temin a Nobel Prize and required a modification of the original scheme for genetic information flow (Figure 3.2).

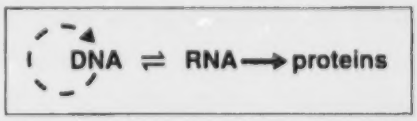


Figure 3.2 Baltimore and Temin's modified scheme for the flow of genetic information.

BEST COPY AVAILABLE

THE CENTRAL DOGMA SPAWNS THE CHICKEN AND THE EGG

A good hypothesis typically yields new questions to be answered; the central dogma was no exception. Biologists interested in the origin and evolution of genetic information mechanisms have long faced a dilemma in trying to figure out which key cellular process, *DNA synthesis* (replication) or *protein synthesis* (translation), was the first to evolve. While trying to answer this question in light of the known biochemistry of contemporary cells and the information flow shown in Figure 3.1, a classic chicken and egg paradox presented itself: If DNA can only be assembled and copied with the aid of protein enzymes, and protein enzymes can only be encoded by a pre-existing DNA, then logically neither one could have arisen first without the other also being present.

THE CHICKEN AND THE EGG MEET THE RNA WORLD

This paradox was apparent to Crick when he proposed his original scheme of information flow. Indeed, he was sufficiently troubled by it that he was one of the first to propose a theoretical solution: Perhaps *RNA* was the first genetic information molecule, emerging before either DNA or protein, and later giving rise to both. This novel proposal remained an interesting speculation for 15 years because, during this time, biologists had no experimental evidence that RNA possessed the biochemical versatility necessary to fulfill the role of progenitor molecule. All of that changed in the early 1980s, when Thomas Cech and Sidney Altman independently discovered the catalytic ability of RNA.

A world of possibilities concerning the origin of life and early evolution opened to biologists once they realized that RNA can potentially fill *both* key roles required of an evolutionary progenitor molecule: encode genetic information and act as a functional catalyst capable of synthesizing other molecules, perhaps DNA and proteins. This new vista is termed the *RNA world hypothesis*, which proposes that RNA originated first and functioned in the earliest stages of molecular evolution as both the encoder of genetic information and the catalytic worker molecule. If correct, this hypothesis eliminates the chicken-egg dilemma (DNA first or protein first) by proposing that RNA came first. But just what functions did these versatile primordial RNAs have that initiated biological evolution at the molecular level? In this activity, you will examine some of the key functions of a primordial RNA.

The RNA world hypothesis expands once again our working scheme of biological information flow and puts it into an evolutionary, historical perspective (Figure 3.3).

Challenge Questions

Read the Introduction, then answer the following questions:

1. Discuss some attributes common to all life forms and generate a brief list of those attributes that you feel are the most essential. Why have you selected these attributes?

2. If the RNA world hypothesis is correct, and RNA was the first heritable biomolecule, then RNA must have used both its information coding and catalytic abilities to accomplish a task essential for its perpetuation. Discuss what this task would be. Is this task included on your list of life's essential attributes?

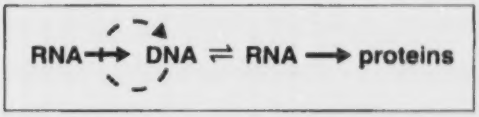


Figure 3.3 The RNA world hypothesis.

3. Modify the information flow scheme shown in Figure 3.3 to reflect your answer to Question 2.

Exercise 3.1: Nucleotide Base Complementarity

In this exercise, you will explore the concept of molecular self-replication. In particular, you will attempt to determine how, very early in the course of biological evolution, a hypothetical RNA molecule might have made copies of itself without the aid of protein enzymes. This property of unaided self-replication is not the rule in living systems today: All DNA and RNA that we know of is replicated by protein enzymes (polymerases or replicases) that assemble the monomer building blocks of these nucleic acids into new copies. Why do we think that RNA molecules might once have been, and may still be, capable of copying themselves?

The rationale stems from the RNA world hypothesis and from the observation that certain catalytic RNA molecules are known to promote the ligation, or joining together, of nucleotide building blocks, while using their own sequence as a template. If a primordial RNA was able to combine information coding and catalytic abilities in this way, it may have been able to make complete copies of itself. This self-replication would have started evolution as the RNA proliferated and inevitably produced variants of itself. Different variants would replicate more or less rapidly, would "compete" with one another for raw materials, and some would eventually assume new functions, such as the synthesis of DNA and proteins.

We are all aware of the universal process of biological reproduction, especially when it results in the creation of a new individual—a plant giving rise to a new plant, or parents giving birth to a child. As biologists, we need to be aware that all *reproduction at the organismal level begins with molecular reproduction—replication—at the level of the genetic molecules DNA and RNA*. At life's most basic level, then, molecular replication is essential.

What are the properties of DNA and RNA molecules that make their replication possible? Quite simply, they are *complementary interaction* and *templating*. The term complementary interaction is used here in the chemical sense: Two molecules interact in a complementary way if their shapes "fit" one another (like a hand in a glove) and they are able to form stable bonding interactions as a result. You know from Activity 1 that in the case of the nucleic acids it is the monomer nucleotide bases A, U (T in DNA), G, and C that engage in such complementary interactions. The interactions follow defined pairing rules, based on the ability of certain base pairs to form stabilizing hydrogen bonds. All known biological systems use this principle of complementarity to replicate their genetic material.

Activity 1 acknowledged the complementary interactions between nucleotide bases and was concerned with them only as determinants of the three-dimensional folding of RNA molecules. Here you will examine the physical interactions of complementary bases more closely and consider the importance of these interactions for the replication of nucleic acids.

BEST COPY AVAILABLE

Materials

- nucleotide base cutouts
- millimeter ruler
- opaque tape

Procedure

1. Arrange paper cutout models of several different monomer nucleotide bases randomly on the desk. Some of the models represent pyrimidine bases, made up of a single ring, and others are purine bases, containing two rings.

2. Group the bases into stable hydrogen-bonded pairs. The dashed lines (- - - -) on the models represent hydrogen bonds. Not all interactions among the bases are stable, but stable interactions will meet the following criteria:

   - An H on one molecule will form a hydrogen bond (- - - -) with an O or an N on the partner molecule.
   - At least two hydrogen bonds between a pair are needed for a stable interaction.

3. List below the stable pairs and the number of hydrogen bonds in each. Also measure the overall diameter of each pair.

Base Pair	Hydrogen Bonds (#)	Diameter (mm)

4. Cover the dashed lines (- - - -) in the uracil (U) with opaque tape and see if you can find another partner for it. Record your findings above.

Challenge Questions

1. How do the stable pairs that you found relate to the purine and pyrimidine base classes?

2. Why are only some base pairings stable? Speculate about the relative stability of pairs with two bonds versus those with three.

3. What pairing(s) did you find for adenine (A)? Which is relevant to DNA? to RNA?

4. What two pairings did you find for uracil (U)? Which of these is a "standard" Watson-Crick pairing? The alternate pairing for U is referred to as a "wobble" base pair. Wobble pairs and a few other nonstandard pairings increase the number and variety of base associations. Predict the general effect of these diverse pairings on the variety of shapes that RNA molecules can potentially assume.

5. How do the relative diameters of stable pairs compare? Assume that the model bases have the same relative geometric proportions as the actual bases; what implication does of your observation on base-pair diameters have for the overall diameter of a double-stranded nucleic acid? How might a G-G mispairing influence the diameter of the double-stranded molecule?

Next you will use the strategy of molecular templating to explore how complementary interactions of bases play a role in the replication of nucleic acid polymers.

You may already be familiar with using a template to make a replica. A sculptor, for example, can employ a template to produce a statue (Figure 3.4): In Step 1, a hollow mold (the template) is filled with liquid material that hardens. Separation of the template from the hardened material yields a product that is *complementary* to the template, in the sense that the product is solid where the template is hollow, and the surface contours of the product protrude where those of the template recede.

Challenge Questions

Refer to Step 2 in Figure 3.4. Assume that the sculptor uses the statue itself, the complementary product, as a new template and pours liquid material over its surface. After the material hardens, the sculptor separates the products.

1. How does the hollow end-product of this second step compare with the objects in Step 1?

2. What is the overall result of this process?

Exercise 3.2a: Discovering a Pathway to Self-Replication

Because our working hypothesis is that a self-replicating RNA is likely to have been the first "living" and evolving molecule, we must test this hypothesis by finding such a molecule in nature or creating one in the laboratory. As yet, no naturally occurring RNA able to fully replicate itself has been found in any organism. Such an RNA could still exist, perhaps in a hot spring or deep ocean vent, as the living fossil that would provide evidence of a past RNA world. Biologists are on the lookout, and time will tell. In the meantime, scientists are developing new experimental approaches that have real potential to discover just such a self-replicating RNA in the laboratory. You will explore this laboratory approach in a later exercise.

In this exercise, you will discover a sequence of events or a mechanism by which self-replication might take place.

Procedure

The following single-stranded RNA sequence represents a hypothetical ribozyme that can act as a generalized ligase able to couple nucleotide monomers and form a

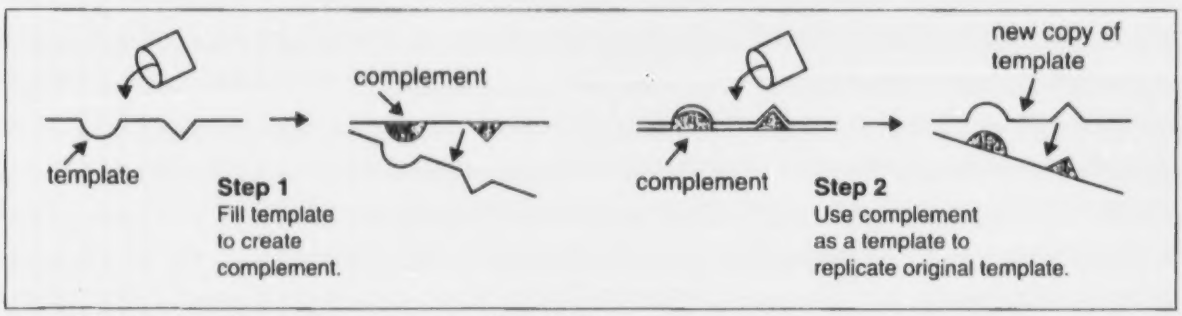


Figure 3.4 Using a template to make an exact copy of an original.

BEST COPY AVAILABLE

new single-stranded RNA molecule. The ribozyme can use either itself or other RNA molecules as a template to guide this assembly. Also shown are supplies of the four nucleotide building blocks, A, U, G, and C. Your goal is to devise a series of steps by which this template/ligase molecule can ultimately assemble an *exact copy* of itself.

1. Apply base pairing and the template strategy while writing in the appropriate monomers next to the template. After completing this step, answer the Challenge Questions that follow.

Template/Ligase Ribozyme	Monomer Pools	
5′		
C		
G	A A A	C C C
U	A A A	C C C
C		
G	G G G	U U U
G	G G G	U U U
A		
A		
A		
C		
G		
U		
A		
U		
C		
C		
3′		

Challenge Questions

1. Which of the following terms applies to the new single-stranded molecule that you assembled?

   a. exact copy
   b. complement

2. Is the molecule that you assembled likely to be able to act as a template? Explain.

3. Is the molecule that you assembled able to also act as a ligase catalyst? Explain.

Procedure (continued)

2. Assume for the moment that the complementary molecule you constructed is a catalytic ligase, like the original template/ligase molecule. How might an exact copy of the original be generated?

3. Now assume a different scenario in which the complement that you initially assembled is *not* catalytic. Proceed again to produce the exact copy of the original template RNA. Hint: Consider where the ligase activity might come from for this next step. Keep in mind the properties of the original template/ligase molecule that you started with.

Challenge Questions

1. How did you solve the puzzle of how to carry out the second step of the self-replication process?

2. A two-step process may be the simplest path to self-replication, but it is unlikely to happen this way. More likely, the process is gradual, with several steps involved. Can you think of another route to self-replication?

Additional Information

Orgel, L.E. (1994, October). The origin of life on the earth. *Scientific American, 271*(4): 77–83.

Joyce, G.F., & Orgel, L.E. (1993). Prospects for understanding the RNA world. In Gesteland, R.F., & Atkins, J.F. (Eds.), *The RNA world* (pp. 1–25). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Although no RNA able to completely replicate itself has yet been found in nature, scientists have made progress toward discovering a self-replicator in the laboratory. These experiments are conducted in the test tube, at the molecular level, and apply the key elements of evolution: *variation*, *selection*, and *replication*. Scientists are learning much about the potential biochemical capabilities of nucleic acids and are developing evidence-based hypotheses about how the first evolving RNAs might have come to be. In Exercise 3.3, you will explore an early and pioneering laboratory approach to the study of molecular evolution.

Exercise 3.3: Molecular Evolution in the Test Tube

In the mid-1960s, researcher Sol Spiegelman developed the first system for studying the replication and evolution of RNA molecules in the test tube (*in vitro*). As his starting material, Spiegelman chose the RNA molecule comprising the genome of bacteriophage Qβ, a virus that normally infects the bacterium *Escherichia coli* (*E. coli*). To be replicated in a natural infection cycle, the Qβ RNA must first get inside an intact *E. coli* cell. To accomplish this, three of the four genes encoded by the RNA's 4,000 nucleotides specify proteins that enable the RNA to enter the bacterial cell and its "progeny" RNAs to spread to new bacteria. The fourth gene encodes viral replicase, the protein enzyme that uses the viral RNA as a template on which to assemble monomers into new copies of the RNA. The replicase enzyme initiates copying of the RNA by first binding to a small subset of bases within it, called the *origin of replication*. These few bases are all that any Qβ RNA molecule needs to be copied by the replicase. Any molecule with an intact origin of replication sequence will be copied, and any molecule in which this sequence is either lost or significantly mutated will not be copied or will be copied at an altered rate.

The experimental system that Spiegelman employed was well suited to the study of molecular replication and evolution *in vitro*:

- The experimental system streamlined and simplified the viral RNA replication process, as compared with a natural *in vivo* infection. Specifically, the system eliminated the requirement for Qβ RNA to first get inside an intact bacterial cell. Spiegelman accomplished this by providing free in the reaction tube all the raw materials needed for RNA synthesis (that is, the A, U, G, and C building block

BEST COPY AVAILABLE

FOR YOUR INFORMATION
Modern Evidence of an Ancient World

As we study the many functions of RNA in contemporary life, we see remnants of early informational and catalytic molecules that could have taken the first unaided steps in building the maze of tiny bridges that led from inanimate chemicals to living creatures. These versatile RNA molecules likely populated a primordial RNA world, in which the process of evolution first began.

The efficient genomes of certain contemporary viruses and even simpler viruslike agents demonstrate that RNA might easily have been a major genetic player in the saga of life. And today's remaining catalytic RNAs (ribozymes) may well be the 4-billion-year-old descendants of early catalysts that made possible key evolutionary processes, such as the replication of RNA itself. Evidence of RNA's early importance mounts as new contemporary RNAs and their diverse functions in the cell continue to be discovered. The familiar mRNAs, rRNAs, and tRNAs involved in gene expression and protein synthesis have been joined by several recently identified molecules, such as the small nuclear RNAs (snRNAs) that are involved in splicing and editing cellular mRNAs. New functions are even being recognized for some of the familiar RNAs. Transfer RNA, for example, has been found to participate in nucleic acid replication and in the synthesis of bacterial cell walls. Some ribonucleotides, the monomer building blocks of RNA, are key components of the coenzymes that assist protein catalysts. Modern coenzymes may be leftovers from an ancient RNA world in which primitive catalysis and metabolism were conducted solely by RNA molecules that eventually acquired the ability to synthesize more efficient proteins. These and many other examples show us that the range of functions for RNA is considerably wider than previously thought (Figure 3.5).

In our search for evidence of RNA's versatility and life's origins, we are not limited to identifying molecular functions that exist in modern cells or viruses. Research scientists now conduct laboratory experiments in which populations of RNA molecules are made to undergo evolution in the test tube, producing molecules with entirely novel structures and functions. These *in vitro* selection experiments demonstrate that populations of RNA evolve according to known principles.

The evidence at hand from several different approaches raises our confidence about ancient scenarios that envision RNA as perhaps the first self-replicating molecule—a pioneer of life.

- RNA stores information and performs catalysis *in vivo*; no other biomolecule has both properties.
- Nucleotide sequences of RNAs common to all organisms (for example, rRNAs) are highly conserved (similar) among the many different species studied, suggesting that RNA was a key molecule present early in evolution.
- RNA or ribonucleotides are involved in most critical cellular functions in all three domains of life:
  - Adenosine triphosphate (ATP) is a universal energy carrier.
  - Universal metabolic pathways employ adenine nucleotide coenzymes (NADH, NADPH, FAD, CoA).
  - Protein synthesis employs mRNAs, rRNAs, and tRNAs.
  - rRNA by itself can catalyze peptide bond formation.
  - DNA synthesis requires the prior conversion of ribonucleotides to their deoxy form.
  - The ribonucleotide uracil, found only in RNA, is the precursor for DNA's thymine.
  - RNA is the primer for DNA replication.
  - Ribonucleotide derivatives function as key signaling molecules in the cell (for example, cAMP, ATP).
- RNAs function as primers in DNA replication and reverse transcription of retroviral genomes.
- tRNA-like molecules are involved in nontranslational polymerizations (for example, cell wall synthesis, antibiotic synthesis).
- A tRNA-like molecule may have given rise to the RNA component of telomerase, the enzyme that maintains the ends of chromosomes.
- Enzymatic processing of mRNAs involves other small RNAs (for example, snRNAs, RNase P).
- Protein sorting into the endoplasmic reticulum of all eukaryotes involves RNA (for example, SRP-RNA).
- During polysaccharide synthesis, ribonucleotides activate and carry sugars.

Figure 3.5 Modern RNA functions that are consistent with an early RNA world.

nucleotides of RNA, plus some accessory bacterial proteins). These materials normally would only be available to Qβ once it is inside a host bacterial cell. Also provided in the tube was an ample supply of the viral replicase enzyme.

- The system also had built-in mutation features ensuring base sequence changes (molecular variation) in the population of "progeny." First, the replicase enzyme is a relatively sloppy worker, making one or two random nucleotide base changes (mutations) in each RNA copy that it produces. Second, the replicase occasionally produces randomly broken copies of the RNA. In a natural infection of *E. coli*, many of these defective, shorter-than-normal RNAs would be uncopyable or unable to spread the infection to other bacteria. Broken molecules in the test tube system, in contrast, were not at a disadvantage for either copying or perpetuation into the next "generation." They could still be copied, provided that *they retained intact the short origin of replication sequence recognized by the replicase enzyme*. And, as described above, their perpetuation was independent of their ability to infect cells in this test tube system.

The test tube system, then, neatly incorporated two of the three features essential for all evolution: replication and variation.

Spiegelman next provided the third essential feature of evolution, selection, to observe the evolution of the starting population of RNA molecules. To apply a *selective pressure* (a condition favoring some individuals and disfavoring others) on the population of molecules, Spiegelman limited the time available to complete replication, thus giving a selective advantage to those RNA molecules that could be copied quickly. In effect, this time limit transformed the copying process into a race between variant RNAs that arose at random during the course of the experiment. Speed of replication became a "phenotypic trait" of the molecules and a test of their molecular "fitness" in this test tube "environment."

Spiegelman and his associate started the experiment by adding Qβ RNA to a test tube containing replicase enzyme and nucleotide monomers. They allowed the

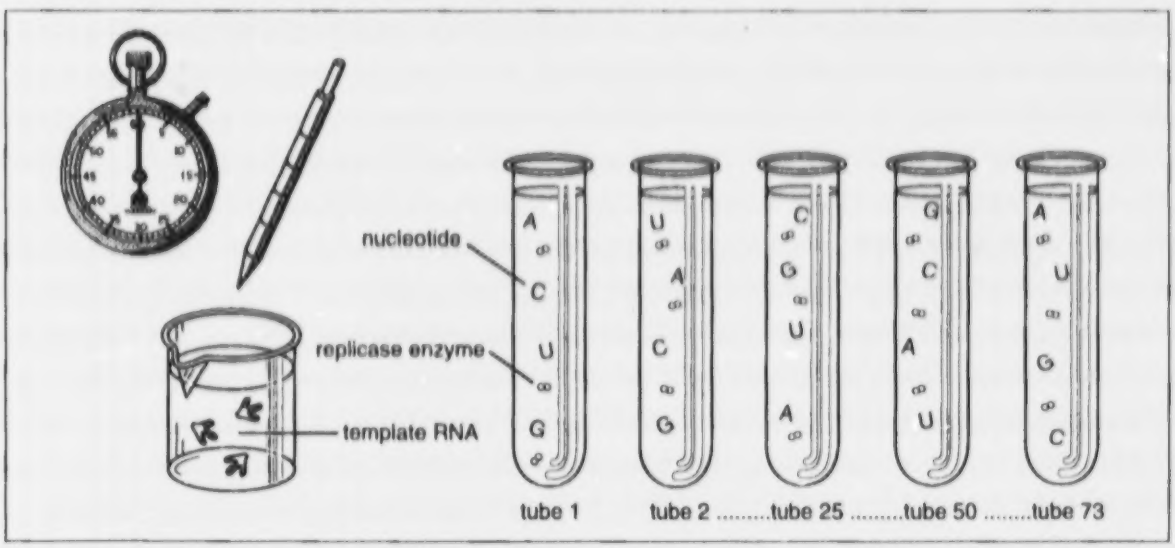


Figure 3.6 Experimental setup for *in vitro* replication of RNA.

BEST COPY AVAILABLE

replication reactions to proceed for just 15 minutes. They then transferred a random sample of progeny RNAs from the first tube into a second tube containing a fresh supply of nucleotides and replicase enzyme (but no RNA other than that was transferred). The replication process was again allowed to proceed for 15 minutes, and a sample from this second tube was transferred to a third tube of fresh raw materials. This serial-transfer process was repeated 73 times (Figure 3.6). During the reaction, the experimenters monitored the total amount of RNA that accumulated in each tube as well as the size and nucleotide base composition of each "generation" of RNA.

Procedure

1. Use what you now know about Spiegelman's experimental system and about evolution to formulate a hypothesis and make general predictions about how the starting RNA population was altered during the course of Spiegelman's experiment. Your hypothesis should make predictions about changes that might occur in three phenotypic traits of the RNA molecules:

   - speed of replication
   - length of the molecule
   - ability to infect *E. coli* bacteria

Challenge Questions

Revise your initial hypothesis if necessary in light of the data in Figure 3.7 and answer the following questions:

1. What can you conclude from Figure 3.7 about the total number of RNA molecules produced in each generation? What does this imply about the average replication speed of the molecules?

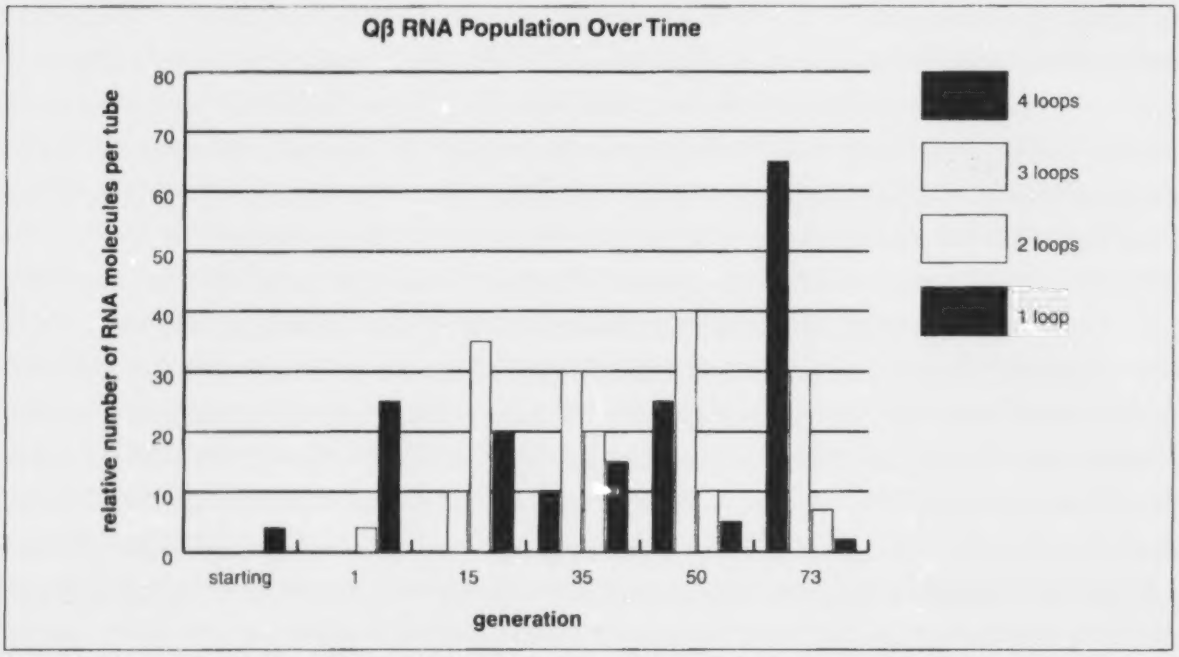


Figure 3.7 Summary graph of simulated results from the *in vitro* evolution experiment.

2. The reason for the change in the speed of RNA replication was

   a. a change in the replicase.
   b. a change in reaction conditions among tubes.
   c. a change in the RNA.

3. Two major changes occurred in the RNA molecules during the course of the experiment that account for their altered replication speed. First, copying errors that replaced one base with another were randomly made by the "sloppy" replicase. The second and major reason for altered replication speed is apparent from the general trend in the size of the molecules across time.

   a. How would you describe this trend?
   b. What happened to the longer molecules?

4. Imposing a time limit on the population for replication effectively made the copying process a competition, a race. This competition was certainly not "intentional" on the part of the molecules, but it was inevitable simply because of the way the experimental "environment" operated. The reason each "generation" of RNA replicated faster than the preceding one is that

   a. the faster molecules in each generation produced more "offspring" copies like themselves, thus increasing their chances of being randomly transferred to the next generation.
   b. each generation became progressively better "adapted" to the demands of the environment.
   c. both are correct.

5. If you were to test the RNAs of successive test tube generations for their ability to carry out natural infection-replication cycles in bacteria, you would expect to find

   a. increased infectivity.
   b. decreased infectivity.
   c. no change in infectivity.

6. After going from a length of 4,000 bases to approximately 700 (shedding more than 80 percent of the genome) the molecules became no shorter. The reason is that

   a. RNAs shorter than this were not sampled and transferred to the next tube.
   b. the replicase could not copy Qβ RNA molecules shorter than this.

7. If you were to test the shortened or mutated RNAs in the test tube for their ability to carry out a natural infection-replication cycle in bacteria, you would expect to find

   a. increased infectivity.
   b. decreased infectivity.
   c. no change in infectivity.

8. A change in environmental conditions will select for new "traits" in molecules, just as it selects for new phenotypic traits in populations of organisms evolving in nature. Spiegelman demonstrated this fact by changing the test tube environment in various ways and repeating the experiment. In one case, he added to each tube a chemical inhibitor of the replication process. He added just enough inhibitor to significantly slow but not completely prevent replication. Over time, the replication speed of the RNAs

   a. decreased.
   b. increased.
   c. remained unchanged.

Since these classic experiments were conducted, powerful new approaches to studying RNA and DNA in the laboratory have been developed. Techniques such as the polymerase chain reaction (PCR) for *in vitro* replication of nucleic acids, rapid nucleotide sequencing methods, and assays of nucleic acid function provide very efficient ways to study these molecules. Scientists are now asking ever more probing questions about nucleic acid functions, those that already exist in nature and those that might be fashioned in the lab. Today, for example, it is possible to synthesize in the test tube large populations of RNA or DNA molecules with randomly varying base sequences. Scientists can select from these diverse populations only those molecules that possess a particular desired function, such as the ability to catalyze a chemical reaction or bind specifically to some other molecule. Nucleic acid sequences discovered in this way might one day be used to disable disease-causing viruses and bacteria or even treat certain forms of cancer.

Directed molecular evolution, as this laboratory approach to molecular evolution is called, is simply a more powerful and efficient variation on Spiegelman's early studies of RNA evolution. The same basic evolutionary features of population variation, selection, and replication are employed in both cases. By exploiting the power of such experimental approaches, we can learn much about the range of potential RNA functions and can use this information to develop evidence-based hypotheses for how the first evolving molecules might have come to be. In Exercise 3.4, you will explore a sophisticated approach to molecular evolution.

Additional Information

Mills, D.R., Peterson, R.L., & Spiegelman, S. (1967, July 15). An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proceedings of the National Academy of Sciences, 58*(1): 217–224.

Joyce, G.F. (1992, December). Directed molecular evolution. *Scientific American, 267*(6): 90–97.

Szostak, J.W., & Ellington, A.D. (1993). *In vitro* selection of functional RNA sequences. In Gesteland, R.F., & Atkins, J.F. (Eds.), *The RNA world* (pp. 511–533). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

FOR YOUR INFORMATION
Molecular Selection in Real Time

One of the great difficulties of studying natural selection in the wild is that selection generally takes place across a long period of time. Often scientists must be satisfied with reconstructing earlier events by looking at the results of those events. For example, in the 19th century, Charles Darwin observed a wide range of beak sizes and shapes among various species of songbirds isolated in the Galápagos Islands. The birds all turned out to be finches, yet their diets and habitats were very different, and their beak characteristics correlated to their different ways to get food. Darwin observed these results of natural selective pressure, but he did not actually watch the process happen.

Fortunately, scientists have had opportunities to observe natural selection taking place in the wild and in real time. Darwin's finches provided such an opportunity. Several scientists extended observations of these finches in the 1970s and later. These researchers collected data during several seasons, measuring and recording the average size of beaks in populations of different finch species and the size and availability of the seeds on which these birds fed. The data showed that during the observation period, changes in average beak size occurred in correlation with changes in the food supply of seeds. Dramatically, changes in the characteristics of the finch population could be observed in just one or two seasons.

Recently, scientists carried out experiments with selection in the laboratory, using populations of molecules. These modern *in vitro* selection (or directed molecular evolution) experiments trace their origins to the RNA replication work of Spiegelman. But they are much more powerful, able to generate RNA or DNA molecules with *predictable* shapes and functions. These experiments are based on the following goal: to direct the evolution of a random population of RNA molecules such that the population becomes enriched for molecules with a desired function.

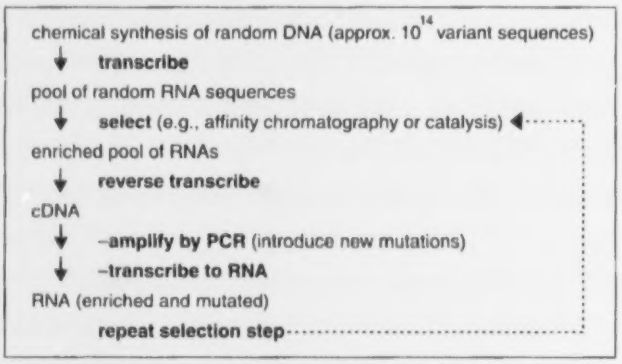

chemical synthesis of random DNA (approx. 10^{14} variant sequences)
transcribe
pool of random RNA sequences
select (e.g., affinity chromatography or catalysis)
enriched pool of RNAs
reverse transcribe
cDNA
–amplify by PCR (introduce new mutations)
–transcribe to RNA
RNA (enriched and mutated)
repeat selection step

Figure 3.8 The strategy of directed molecular evolution. A large, randomly synthesized population of nucleic acids undergoes repeated cycles of selection followed by amplification. The result is a particular sequence well suited to the selection criterion.

For directed molecular evolution experiments (Figure 3.8), scientists start with a large population of random RNA molecules (approximately 10^{15}), each with its own particular base sequence and shape. This starting mixture provides the variation that is required for evolution to take place. The scientists subject this diverse population of molecules to a selective test (also called *selective pressure*) by requiring that molecules possess some specific function, such as binding to a test substance or catalyzing a chemical reaction, before the RNA molecule can be replicated. Molecules able to meet the selective pressure have a reproductive advantage over the others. Selective reproduction (replication) of RNAs passing the test is typically accomplished by the laboratory procedure known as the *polymerase chain reaction* (PCR), in which many copies of the RNA molecules can be generated in the test tube. By repeating this cycle of selection and reproduction of successful variants several times, the final population of RNA molecules is no longer random. Instead it is enriched for particular structures that can carry out the desired molecular function.

FOR YOUR INFORMATION (continued)

Directed molecular evolution is simply a more powerful and efficient variation on Spiegelman's pioneering *in vitro* studies of RNA evolution. Both of these laboratory experiments are nevertheless true examples of evolution at the molecular level. The same basic elements are at play here as in the evolution of a population of living organisms: *variation in the population, selection of individuals based on some essential function or ability,* and *selective reproduction of these individuals.*

In vitro molecular evolution is, of course, much faster than the evolution of complex organisms. But both processes are rather inefficient: Many molecules or organisms must be tested for each one able to successfully meet the test. Nevertheless, the evolutionary process for both cases is powerful across time, being able to generate RNA molecules with new functions and populations of organisms adapted to changing environments.

Ideally, an *in vitro* process would be able to predict the exact RNA base sequence and shape required for a given task and synthesize only this molecule, rather than having to screen many billions of random molecules. Unfortunately, we are a long way from knowing enough about how molecular composition determines shape and function to be able to use this more efficient approach. Nevertheless, as we learn more from the *in vitro* selection experiments about the capabilities of RNA and DNA, we come closer to this goal.

Exercise 3.4: A Nifty Trick with RNA in the Laboratory

A prediction of the RNA world hypothesis is that a self-replicating RNA once existed in nature (and may still exist) and initiated the process of biological evolution. In Exercise 3.2, you saw how, in principle, RNA might employ its catalytic ability to replicate itself, and you examined laboratory approaches, like directed molecular evolution, that can evolve RNAs with specific abilities. The search for a self-replicating RNA continues in the lab and in the field. In this exercise, you will focus on a recent laboratory product of this search, an RNA molecule that neatly unites the ideas of directed molecular evolution, RNA catalysis, and RNA replication.

Procedure

Figure 3.9 depicts an experimental reaction system for the replication of RNA. This system was developed recently in the lab of Gerald Joyce at The Scripps Research Institute. The RNA molecule at the heart of this system functions as both a template molecule and a catalyst. It was originally generated in the laboratory using the technique of directed molecular evolution by screening a large population of 10^{15} randomly synthesized RNA molecules. The researchers were able to select from this varied population an RNA sequence, designated E100, capable of acting efficiently as both template and ligase catalyst in this system.

To understand what is happening in Figure 3.9, apply the templating strategy for replication that you studied in Exercise 3.1 (Figure 3.4). Recall that just as in sculpture, the strategy to reproduce an original object is to make the complement of the original, and then make the complement of that complement: *The complement of the original's complement is the original.* As you work through this reaction sequence

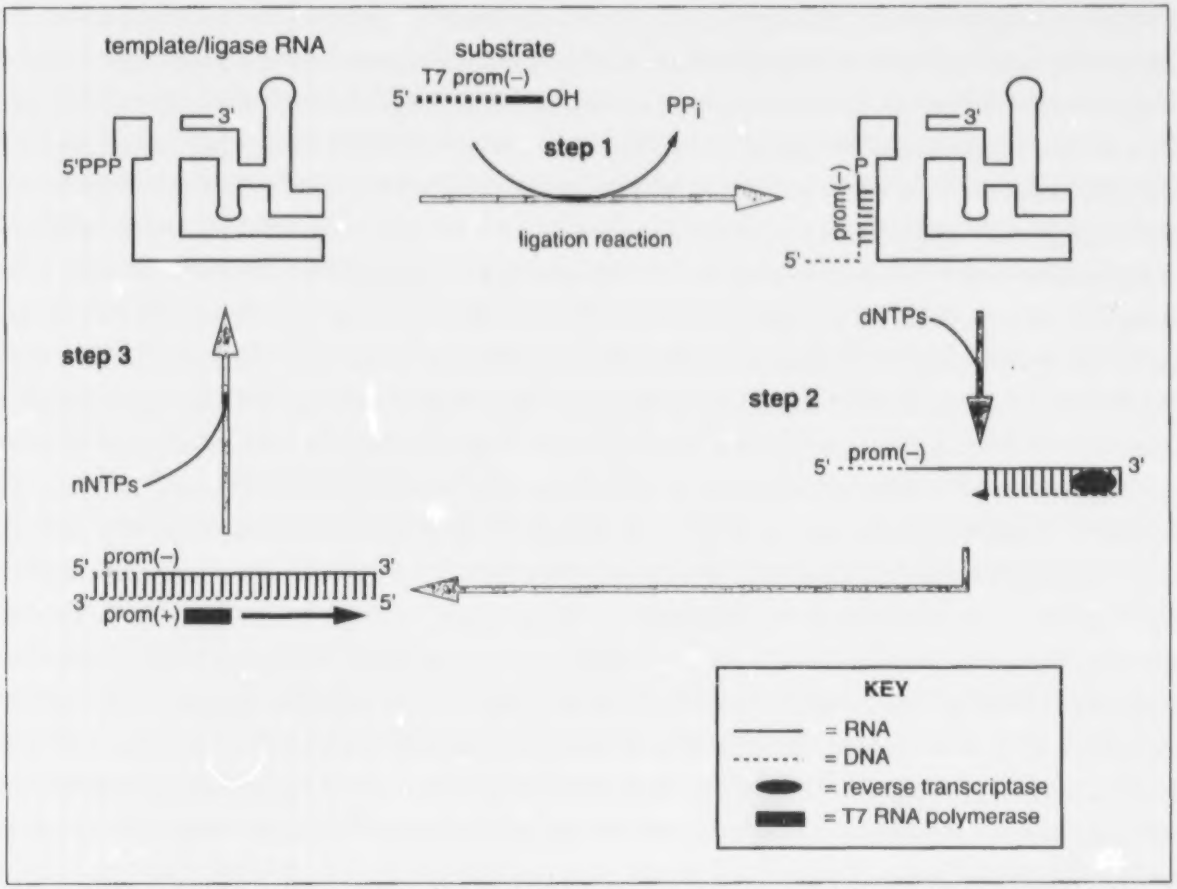


Figure 3.9 *In vitro* system for the continuous replication of RNA. RNA strands are shown as solid lines, DNA strands as dashed lines.

and make sense of the individual steps, you will appreciate its elegant internal logic as a simple application of the templating principle. In a later exercise, you will have the chance to perform this replication reaction in the laboratory.

Begin this exercise by analyzing what is happening in Figure 3.9, then answer the Challenge Questions.

Step 1

- The RNA molecule in the upper left of Figure 3.9 is the ribozyme that serves both a template function and a ligase catalyst function in this system. We will refer to it as the *template/ligase*.
- The molecule labeled *T7 prom(-)* is a short single strand made up of DNA nucleotides, with a few RNA nucleotides at its 3'-end. T7 prom(-) is the substrate for the template/ligase, and it encodes one strand of the T7 promotor. Later in the reaction, this promotor will serve, in double-stranded form, as the start site for a transcription event by the protein enzyme T7 RNA polymerase. (Recall that RNA polymerases, in general, transcribe DNA templates to make RNA molecules.)

- In Step 1, the template/ligase uses its template ability to properly align itself with complementary bases in the substrate; it then uses its catalytic ability to ligate the substrate to its own 5′-end.

Step 2

- The DNA primer is a short sequence of DNA that is complementary to the 3′-end of the template/ligase RNA (shown hybridized in Step 2; the RNA is shown unfolded at this point for simplicity). The primer's sole function in the reaction is to provide a start site for the protein enzyme reverse transcriptase. Recall that reverse transcriptase copies an RNA template (or a DNA template) into a DNA strand.
- The result of the reverse transcriptase step in this reaction is a double-stranded molecule that is a hybrid. One strand of the hybrid is the original ribozyme RNA with its attached prom(-) portion of the T7 promotor; the newly made second strand is DNA that is complementary to the ribozyme (and also provides the second strand of the T7 promotor).

Step 3

- T7 RNA polymerase uses the now-functional (double-stranded) T7 promotor as its start site to carry out a transcription reaction.

Challenge Questions

1. What is accomplished in Step 1 of the reaction sequence? Which function(s) of the RNA ribozyme are employed at this step?

2. A double-stranded molecule results from Step 2, reverse transcription. What is the coding relationship of the newly synthesized second strand to the RNA ribozyme sequence? Which part of the templating strategy is accomplished in Step 2? At the completion of this step, what function can the T7 promotor in the double-stranded molecule perform?

3. In terms of the templating strategy, what does RNA polymerase accomplish in Step 3? What is the product of this step? What effect does this step have on the number of template/ligase RNA molecules in the tube?

4. Assume there is a surplus of monomer building blocks (dNTPs and rNTPs) and substrate molecules in the tube. What will happen across time? If an aliquot of this reaction mixture is transferred to a new tube containing monomers and substrate molecules, what would you expect to happen?

5. The enzymes employed in the system, especially RNA polymerase, have relatively high error rates (approximately one to two errors per molecule copied). Comment on this system's potential to generate novel RNA sequences in the laboratory.

6. The Joyce system, like the Qβ RNA replication system developed by Spiegelman, is able to reproduce RNAs continuously. Both of these systems, however, employ reaction components that prevent them from qualifying as the true self-replication of RNA. What are they? Which system is closer to true self-replication? Why?

Exercise 3.5: Continuous *in vitro* Replication of RNA

Here is your opportunity to carry out in the lab the RNA replication reactions that you explored in Exercise 3.4. Your teacher will provide you with the following reagents:

REAGENT LIST

a. Tube 1, stock mix (This tube will be prepared ahead of time for you.)

218.2 µl high quality distilled water	
20 µl KCl (1 M)	final concentration during reaction = 50 mM
12 µl EPPS buffer (1 M, pH 8.5 at 22°C)	30 mM
10 µl $MgCl_2$ (1 M)	25 mM
8 µl Spermidine (100 mM)	2 mM
20 µl dithiothreitol (DTT) (100 mM)	5 mM
8.4 µl cDNA primer, TAS 1.23 (100 µM)	2 µM
32 µl ribonucleoside triphosphate (rNTP) mix* (25 mM of each rNTP)	4 x 2 mM
3.2 µl deoxy-NTP (dNTP) mix** (25 mM of each dNTP)	4 x 200 mM

* rNTP mix is prepared ahead by adding an equal volume of each of the four individual rNTPs (each at 100 mM) to a single tube.

** dNTP mix is prepared by adding an equal volume of each of the four dNTPs (each at 100 mM) to a single test tube.

b. Tube 2, 22 µl substrate (S162-2) (100 µM)

c. Tube 3, 35 µl T7 RNA polymerase (100 U/µl)

d. Tube 4, 35 µl M-MLV reverse transcriptase (200 U/µl)

e. Tube 5, 2.6 µl oxazole yellow (YO-PRO-1) dye (1 mM). Protect from light.

f. Tube 6, 54 µl input PCR DNA (16 nM)

Procedure

You will use a dye that fluoresces when bound to the nucleic acid to visualize the production of RNA across time.

CAUTION: Always wear eye protection when viewing.

1. Place tube 1, tube 6, and a spectrofluorimeter cuvette (0.5 ml capacity) at 37°C.

2. To tube 1 at 37°C, add the following, in order:

   16 µl of T7 RNA polymerase from tube 3
   16 µl of M-MLV reverse transcriptase from tube 4
   10 ul substrate (S162-2) from tube 2
   1.2 µl of oxazole yellow dye from tube 5

BEST COPY AVAILABLE

3. To tube 1, add 25 µl (400 fmol) of prewarmed input PCR DNA (from tube 6).

4. Immediately transfer this mix to the prewarmed cuvette and begin taking spectrofluorimeter readings (excitation wavelength = 491 nm; emission wavelength = 509 nm).

5. Run the reaction in the cuvette at 37°C for 1 hour, taking fluorescence readings every 2 minutes. As you monitor the reaction, remind yourself of what is occurring in the tube by reviewing Figure 3.9 in Exercise 3.4.

6. Plot fluorescence readings versus time to observe the RNA growth curve.

Challenge Questions

1. How did the level of fluorescence emitted by the reaction change across time? How do you account for this change?

2. How might this system of continuous replication be applied in a directed molecular evolution type of experiment?

Alternative Exercise 3.5: Continuous *in vitro* Replication of RNA Using Visual Detection of Fluorescence

Here is your opportunity to carry out the RNA replication reactions that you explored in Exercise 3.4. Your teacher will provide you with the following reagents:

REAGENT LIST

a. Stock mix

16 µl KCl (1 M) final concentration during reaction =	50 mM
9.6 µl EPPS buffer (1 M, pH 8.5 at 22°C)	30 mM
8 µl $MgCl_2$ (1 M)	25 mM
6.5 µl Spermidine (100 µM)	2 mM
16 µl dithiothreitol (DTT) (100 mM)	5 mM
26 µl nucleoside triphosphate (NTP) mix (25 mM of each NTP)	4 x 2 mM
2.6 µl deoxy-NTP (dNTP) mix (25 mM of each dNTP)	4 x 200 µM
6.5 µl cDNA primer - TAS 1.23 (100 µM)	2 µM
229.2 µl high purity distilled water	
320.4 µl total volume	

b. Tube 1, dried 20 µl aliquot of stock mix

Each group of students will receive *two* each of tube 1. Each group will also receive *one* each of tubes 2–6; each of these tubes has 2X volume of reagent, enough to replicate the experiment twice.

c. Tube 2, 5 µl substrate (S162-2) (25 µM)

d. Tube 3, 4 µl T7 RNA polymerase (80 U/µl)

e. Tube 4, 4 µl M-MLV reverse transcriptase (200 U/µl)

BEST COPY AVAILABLE

f. Tube 5, 160 µl oxazole yellow dye (YO-PRO-1) (5 µM). Protect from light.

g. Tube 6, 24 µl input PCR DNA (16 nM)

Procedure

You will use a dye that fluoresces when bound to the nucleic acid and illuminated with UV light to visualize the increase in RNA across time.

CAUTION: Always wear eye protection when viewing.

1. Label six half-milliliter Eppendorff tubes as follows: 0, 5, 10, 15, 20, and 25 minutes.
2. Put the tubes on ice; to each, add 13 µl of oxazole yellow dye from tube 5.
3. Place tubes 1 (dried stock mix) and 6 (input PCR DNA) at 37°C.
4. Add the following to tube 1 at 37°C:
   - 5.4 µl distilled water
   - 2.0 µl of substrate (S162-2) from tube 2
   - 1.3 µl of T7 RNA polymerase from tube 3
   - 1.3 µl M-MTV reverse transcriptase from tube 4
5. At time zero, add 10 µl of prewarmed input PCR DNA to tube 1. Mix and immediately remove a 2 µl aliquot from this tube and transfer it to the iced dye tube labeled "0."
6. At 5-minute intervals, transfer a 2 aliquot from tube 1 to the appropriately labeled dye tube.
7. Visualize the six dye-containing tubes as a group by UV light or, ideally, 490 nm light.

Challenge Questions

1. How did the level of fluorescence emitted by the reaction change across time? How do you account for this change?
2. How could this system of continuous replication be applied in a directed molecular evolution type of experiment?

FOR YOUR INFORMATION
What Airplane Design Hasn't Got to Do with Biology

When we look at the effects of natural selection on a population of organisms or of molecules, we see that certain traits become more widespread in the population if they provide some advantage under a particular set of circumstances. This phenomenon is commonly known as survival of the fittest, and it is easily observable. It is tempting, then, to think that whatever characteristic or individual comes to dominate a population must be the best structure possible to carry out the needed functions. This view is in error. When we look at a living organism or even at individual molecular structures, we realize that an engineer would have designed them differently.

Many people are under the impression that machines evolve just like living organisms, but this is hardly the case. Consider a simple example, the human-designed airplane. The Wright Brothers achieved the first successful powered air flight by building a biplane. It had a wooden frame, a cloth skin, an open cockpit, and a propeller driven by a reciprocating piston engine. This frail craft would not be much competition for a large modern passenger plane carrying 300 times as many passengers and flying 100 times faster at much higher altitudes.

Although we may say casually that one plane "evolved" into the other, in fact, what really happened over the years was that many new prototype planes were created. But each new plane was designed, tested, and built on the ground before it was flown. Understandably, engineers choose not to take the risks associated with construction on-the-fly. Living systems do not have this luxury.

A living system has to keep going, keep reproducing, without missing a generation, or it disappears. Cells do not get to take time out while the machinery for protein synthesis or replication is replaced by a completely new and better design. This would be analogous to the old parts of an airplane being dismantled in midair while new parts of better design were constructed and attached without ever landing or crashing. Intermediate structures, such as a plane with one wooden wing and one new metal one, would have to be sufficiently airworthy for uninterrupted flight.

Because life comes from pre-existing life, living organisms must in effect evolve in "midair." The fictional Dr. Frankenstein went into the laboratory and constructed a new human from spare parts before he sparked it into life, but in the real world of living systems, life is continuous. New organisms must be built from the blueprints of their parents without interruption.

That biological evolution must take place on-the-fly is an enormous constraint. The feasible alterations to an airplane in flight are much more limited than the experiments and changes that engineers can carry out on the ground, in the safety of the hangar. The engineer builds painstakingly with planning and foresight; evolution meanders along testing randomly generated variations. The combined effects of trial and error plus natural selection take an unavoidable toll in failed experiments, but inevitably lead to better designs for the survivors. Importantly, what survives in nature will not necessarily be the best design possible but simply the most advantageous structure available at the time from a large pool of variants.

Activity 4
RNA Evolution in Health and Disease

Introduction

A news article carried the following ominous headline:

> "Overprescribing: Misuse of Antibiotics Creates Superbugs"
> *The Salt Lake Tribune*
> September 17, 1997

The article went on to report:

> Faced with patients demanding medicine for coughs, congestion and sniffles, doctors wrote 12 million [inappropriate] prescriptions for antibiotics to U.S. adults in a single year—even though the drugs are worthless for colds and other viral infections. Such misuse fuels the spread of bacteria that are resistant to antibiotics, leaving fewer effective drugs for patients with serious bacterial infections, University of Utah physician Merle A. Sande and Colorado doctors reported today in *The Journal of the American Medical Association*. "It's extremely serious," said Sande, the university's chairman of internal medicine. "We are losing all these antibiotics. Our future and our children's future is going to depend on us being more selective."

No doubt you have encountered similar news stories warning of the overuse of antibiotics and its contribution to the emergence of resistant strains of bacteria. Since 1943, when penicillin was first introduced as the "magic bullet" for curing many infectious diseases, more than 100 additional antibiotics have been developed. Despite this seemingly large arsenal, the unfortunate fact is that most of these agents are becoming less and less effective as widespread misuse promotes the development of resistance among bacterial species. Today, it is estimated that 90 percent of all staphylococcus strains are penicillin-resistant, and several other pathogenic species, such as *Streptococcus*, *M. tuberculosis*, and *P. aeurginosa*, have developed strains resistant to all but a few remaining drugs. Inappropriate use of antibacterial agents is a problem not just in medical practice but in animal husbandry, where low levels of antibiotics are routinely included in livestock feed, and increasingly in consumer products such as soaps and detergents that incorporate antibacterial agents.

In this activity, you will see that the acquisition of antibiotic resistance by bacteria, and of resistance to antiviral drugs by viruses, is a predictable result of the evolution of these organisms' genetic systems.

BEST COPY AVAILABLE

When you examined RNA replication in Activity 3, you saw that evolution, whether of populations of molecules in the test tube or organisms in the wild, is driven by three basic processes: (1) random generation of mutations in DNA or RNA; (2) replication of these mutations during nucleic acid synthesis, with some of the offspring inheriting these mutations; and (3) enhanced reproductive success (by natural or artificial selection) of those individuals carrying mutations that are advantageous. Now you will explore how these three natural processes can conspire to promote the spread of resistance among pathogens.

Exercise 4.1: Setting the Stage for Antibiotic Resistance

Random mutations are always arising within populations. Most mutations are harmful to their carrier, some are advantageous, and others are "selectively neutral," having no effect on the carrier's reproductive success. Although mutations are inevitable, most organisms manage to keep the *mutation frequency* (the number of mutations per base per generation) relatively low by repairing most of them before they have a chance to be passed on. Thus, the chance of a *particular* mutation occurring in a *particular* individual is low. However, mutations appear often in rapidly growing populations consisting of many individuals.

Consider the bacterial mutations that result in resistance to the aminoglycoside antibiotic streptomycin. Aminoglycoside antibiotics inhibit protein synthesis in prokaryotes (bacteria) by binding to a specific sequence of bases in their 16S ribosomal RNA (see FYI essay *How Aminoglycoside Antibiotics Target Bacterial RNA*). Particular single-base (or point) mutations in the bacterial rRNA can prevent the binding of streptomycin, enabling the bacterium to become resistant to the drug. The likelihood that any particular base will be mutated to cause resistance is small, occurring in perhaps 1 in 1 billion genome replications (that is, the mutation frequency is 1/1,000,000,000 or 10^{-9} per individual). But because bacteria reproduce so frequently (every 20–30 minutes), an infected wound may easily contain billions of individuals. The population in the wound can become large enough that it is virtually certain that at least one bacterium will have acquired streptomycin resistance due to this mutation.

To calculate the probability of finding a particular point mutation in a population of a given size, we would multiply the mutation frequency per base pair per organism (10^{-9}) by the number of organisms present. For example, in a population of 1 billion (10^9) bacteria, the probability equals unity ($10^{-9} \times 10^9 = 1$). That is, the presence of one streptomycin-resistant bacterium is statistically a near certainty by the time the population reaches 1 billion. While one resistant bacterium in a billion might seem like a harmless drop in the bucket, this solitary resistant cell (unlike its millions of susceptible neighbors) would continue to multiply in the presence of streptomycin, producing resistant descendants. Once this mutation has spread in the bacterial population, treatment with streptomycin and certain other related compounds would prove futile.

Another important factor affecting the rate at which bacteria acquire resistance mutations is the fact that bacteria are quite promiscuous. They have several ways of acquiring and exchanging genetic information, with other members of their own species as well as with other bacterial species. These genetic exchange mechanisms include sexual transfer (conjugation), direct uptake of free-floating DNA (transformation), and

transfer from viruses (transduction). Such promiscuous genetic exchange allows for the "accelerated evolution" of resistance in bacteria. (The evolution is accelerated in the sense that a given species can bypass the mutation process by acquiring an already mutated resistance gene from another species.)

Challenge Questions

Read the introduction to Exercise 4.1, then answer the following questions:

1. Infecting bacteria typically divide once every 30 minutes. If a single bacterium is introduced into a wound at time zero, how many hours would it take for this bacterium to exceed 100 descendants? 1,000 descendants? 10,000 descendants?

2. Is it possible that a bacterial population size could reach 1 billion *without* a streptomycin-resistant individual arising? Could it reach 3 billion? Explain.

3a. Is it more likely or less likely that the same mutation would arise independently in two different bacterial cells within a population? Explain.

3b. In reality, there are several sites within the bacterial rRNA that can undergo single point mutation and can individually cause resistance to streptomycin. Each of these sites has a roughly equal likelihood of incurring a mutation (approximately 10^{-9}). What effect does the presence of several potential sites for mutation have on the overall chance of streptomycin resistance arising in the bacterial population?

4. When antibiotics are required to combat an infection, it makes sense to start therapy as soon as possible to cure the disease quickly and minimize suffering. Explain the evolutionary rationale for starting therapy earlier rather than later.

Review Figure 4.1, then answer the following questions:

5. What can you hypothesize from Figure 4.1 about the type of chemical interaction that contributes to the binding between streptomycin and the rRNA?

6. Notice in Figure 4.1b that bases A532 and C525 are unpaired and located in the loop region of the binding stem-loop. Speculate how a mutation from A to G at position 532 could alter the binding of streptomycin and promote resistance.

7. The "530 stem-loop" in 16S rRNA is highly conserved among bacteria, varying very little in sequence between bacterial species.

   a. What does this fact imply about the importance of this region in normal bacterial ribosomal function? Explain.

   b. Speculate why the ability to produce agents like streptomycin was strongly selected for among certain fungi (genus *Streptomyces*) that compete with both bacteria and other fungal species in the environment.

   c. Speculate why streptomycin has no effect on eukaryotic protein synthesis.

FOR YOUR INFORMATION
How Aminoglycoside Antibiotics Target Bacterial RNA

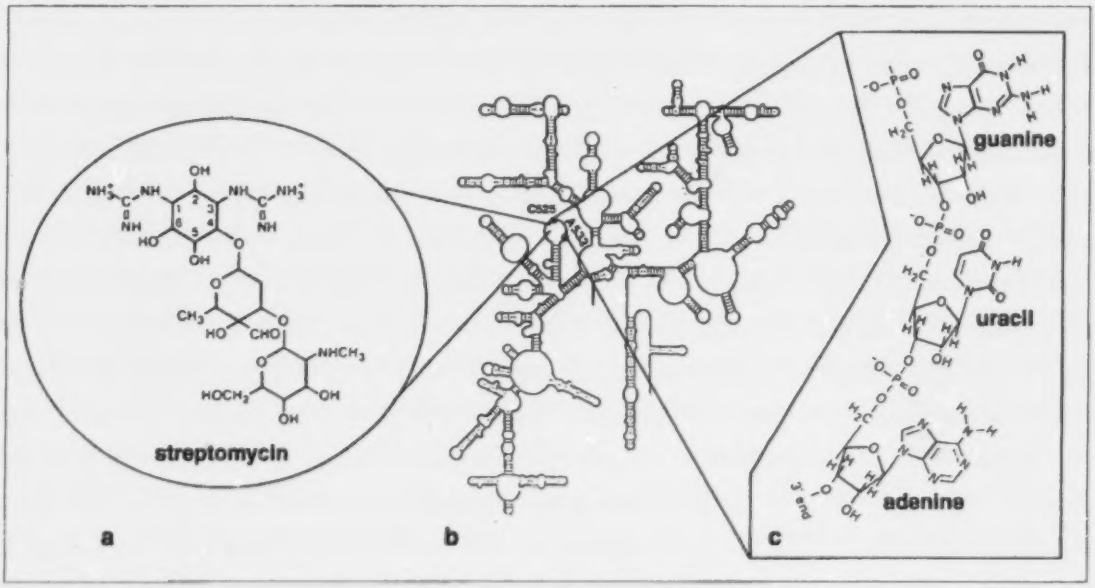


Figure 4.1 (a) The structure of streptomycin. Note the positive charges. (b) The predicted secondary structure of bacterial 16S rRNA, showing the location of the "530 stem-loop" involved in streptomycin binding. The positions of two bases—A532 and C525, whose mutation can confer streptomycin resistance—are indicated. Note that the RNA is shown at smaller scale than the streptomycin molecule. (c) An enlargement of three rRNA bases in the streptomycin binding region, including A532. Note the negative charges in the phosphate backbone.

Exercise 4.2: Selecting for the Emergence of Resistance

The ideal effective antibiotic therapy kills all of the infecting organisms before their rapid division can give rise to even a single highly resistant cell. Unfortunately, this ideal is undermined when the invading population is exposed to a dose of the drug that is too low. The trick is for the patient to start therapy as soon as possible and take enough medication across several days to allow tissue levels of the drug to reach a concentration capable of killing all members of the population.

But the ever present genetic diversity in biological populations complicates this task. Recall that for the vast majority of genetic traits, natural populations have a diversity of phenotypes that display the traits to greater or lesser extent. This is also true for the trait of drug resistance in bacteria. Genetic diversity for antibiotic sensitivity within a bacterial population results in individuals with susceptibilities that range from high to low; greater concentrations of the drug are required to kill the less susceptible individuals.

If too little antibiotic is prescribed or if the patient stops taking it before completing the treatment course, tissue levels of the drug never become high enough to kill all the bacteria. Low levels of streptomycin, for example, merely slow the growth of

bacteria but do not kill them. Such incomplete treatment is an opportunity for bacteria to evolve into a more drug-resistant population. Under these conditions, the antibiotic is said to exert selection pressure on the population. That is, the antibiotic acts as an environmental factor that allows a formerly uncommon phenotype (less susceptible) to multiply while inhibiting the previously dominant phenotype (susceptible). Low levels of the drug may select *against* the majority of the population, which is highly susceptible, by inhibiting or killing it. But these low levels end up selecting *for* the minority of the cells, which are less susceptible, by failing to kill it. These less susceptible bacteria continue to divide and make up an ever increasing portion of the evolving population.

Any new random mutation in this modified population that happens to produce an even higher level of resistance will quickly spread in the same way. The inevitable outcome is a population of highly resistant descendants. Further treatment with this drug will now be ineffective, even at higher doses. The only hope at this point is to try another antibiotic, a diminishing option as organisms continue to evolve resistance to ever more agents.

Inappropriate antibiotic treatment promotes the spread of resistance in pathogen populations in another way: It inhibits or kills normal resident bacterial cells that happen to be sensitive to the drug used. This treatment changes the ecology of the wound "environment" and effectively decreases the competition between bacterial species, which would otherwise act to slow the growth of the pathogen. Making matters worse, it is possible that the resistant pathogen can transfer its resistance gene(s) to other bacterial species, both normal flora and other potential pathogens.

Challenge Questions

Read the introductory text to Exercise 4.2, then answer the following questions:

1. Explain the evolutionary justification for the following statement: The routine addition of antibiotics to livestock feed is ill advised. Can you think of a recent example of a resistant pathogenic bacterial strain that originated in livestock?

2. A common medical practice is to treat single-pathogen infections with broad-spectrum antibiotics that affect several different bacterial species, rather than use a drug that targets particular pathogens. Comment on this practice from the perspective of an evolutionary biologist.

3. Can the emergence of antibiotic resistance be avoided altogether? Explain.

Exercise 4.3: Observing Evolution in Action

In Activity 3, you explored some experimental approaches, such a directed evolution, that allow researchers to manipulate the evolution of populations of RNA molecules in the laboratory. In this exercise, you apply selection to shape the evolution of a population of living cells.

You will observe the effect of varying concentrations of streptomycin on a population of the bacterium *Escherichia coli (E. coli)*. You will accomplish this by growing

BEST COPY AVAILABLE

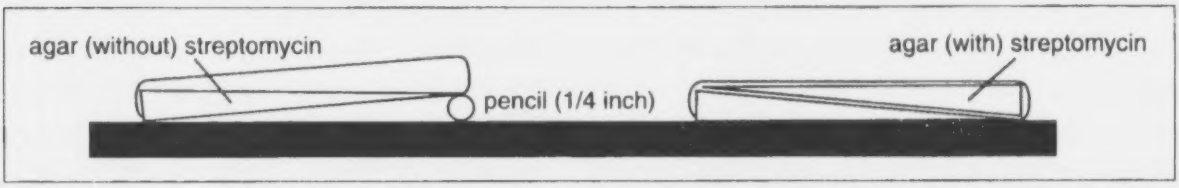


Figure 4.2 Preparation of an antibiotic gradient plate. (1) Pour the lower layer of nutrient agar, containing no antibiotic, at an angle and allow it to solidify. (2) Pour the upper layer of nutrient agar, containing 1 μg/ml of streptomycin, while the plate is level.

the bacterium in the presence of a concentration gradient of the antibiotic. You can easily establish a gradient on a single petri dish using the gradient plate technique (Figure 4.2). Across time, diffusion of the streptomycin from the upper layer into the lower layer establishes a concentration gradient of the drug (from low to high), running from one side of the plate to the other. Cells growing at different locations on the plate are thus subjected to different concentrations of the drug.

After the bacterial population has grown in the streptomycin gradient, you will observe the distribution of growth across the plate and draw conclusions about the population.

Materials

- 1 water bath (for the entire class)
- 1 sterile petri dish
- 1 automatic pipette (0.1–0.2 ml)
- 1 bent glass rod "hockey stick"
- beaker of 70% ethanol
- marking pencil
- 1 tube of starting bacterial population (a 24-hour culture of *E. coli* grown in nutrient broth)
- 2 10-ml tubes of trypticase soy agar
- 1 tube of stock streptomycin solution (10 mg per 100 ml water)

1. Melt the trypticase soy agar in both tubes by immersing them in the hot water bath. When the agar has melted, place the tubes at 45°C.

2. Use the marking pencil to draw an arrow across the middle of the petri dish bottom (on its *outside* surface).

3. Place a pencil under the edge of the plate perpendicular to the arrowhead. Pour in enough molten agar to just cover the entire bottom surface. Allow the agar to solidify in the slanted position.

4. While the first layer is solidifying, use a sterile pipette to add 0.1 ml of the stock streptomycin solution to the second tube of molten agar. Mix well by vortexing or swirling the tube.

5. Place the petri dish in a level position and pour in enough of the streptomycin-containing agar to just cover the high edge of the lower agar layer. Allow to solidify.

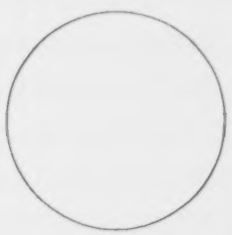
6. With a sterile pipette, deposit a 0.2 ml inoculum of the *E. coli* population on the surface of the agar. Use an alcohol-dipped and flamed bent glass rod to spread the 200 million or so individual cells uniformly over the entire surface of the agar.

7. Label the outside of your plate and incubate in an inverted position for 48 hours at 37°C.

8. Following the first incubation, examine the distribution of bacteria on the plate. You will see areas of confluent growth, where many individuals from the original inoculum are growing crowded together, and areas where discrete colonies are separated by regions showing no growth. Recall that each colony consists entirely of the offspring of a single original cell at that location on the plate, all dividing together. (In a genetic sense, you can think of each colony, with its approximately 10^7 cells, as the clonal equivalent of one cell from the original inoculum).

   a. On the following diagram, draw your arrow and indicate the *high* and *low* ends of the streptomycin gradient.

   b. Indicate on the diagram the locations of confluent growth and of discrete colonies.



   Then select two or three isolated colonies in the middle region of the plate and, using a sterile inoculating loop, streak each colony toward the *high* streptomycin side of the plate.

9. Reincubate the plate in an inverted position for an additional 48 hours at 37°C.

10. Following the second incubation, indicate on the following diagram the pattern of growth that resulted from the colonies that you streaked toward the high streptomycin concentration.

Challenge Questions

1. What can you conclude about those cells in your original *E. coli* population that were able to grow in the higher concentrations of streptomycin, even though they had never been exposed to the drug before you plated them? How does this result relate to the genetic diversity that existed within the original population of bacteria used in the experiment?

2. Assume that you take a few cells from one of the colonies growing at high streptomycin concentration and grow a population of cells from them. How would this population differ from the original one in terms of its proportion of streptomycin-resistant individuals? How does the emergence of this new population represent the workings of selection and evolution?

3. In your experiment, the levels of streptomycin obviously were not high enough to kill all the bacteria in the population. State the parallels between the results of your experiment and the emergence of resistance among the infecting bacteria in patients treated with inadequate doses of antibiotic.

4. Recall that streptomycin inhibits bacterial growth by binding to the organism's 16S ribosomal RNA. Bacteria become resistant to streptomycin primarily as a result of mutations in the 16S rRNA or in one of the ribosomal proteins normally associated with this RNA. Can you determine from the results of this experiment whether the resistance in your bacterial population is due to a change in the rRNA or in the ribosomal protein? Review Figure 4.1 and try to think of an answer that uses the technique of nucleic acid sequencing.

Exercise 4.4: Evolution in a Microcosm

In this exercise, you will see how an RNA virus such as HIV can evolve even within the microcosm of a single infected individual.

Figure 4.3 lists the variant strains of HIV that emerged and predominated at different times during the course of a particular patient's treatment regimen. Nucleotide base sequences were determined in samples of viral RNA (or proviral DNA) taken from the patient at various times before, during, and after treatment with the antiviral drug zidovudine (AZT) alone. AZT is a commonly used component of antiviral drug cocktails because it can inhibit reverse transcriptase (RT). RT is an important, virally encoded enzyme responsible for replicating the virus's RNA and copying it into a DNA provirus form. The DNA provirus can then integrate into the infected cell's genome to produce a state of chronic infection.

	Time of Sample During Treatment (Weeks)						
	Before Treatment	17	22	56	81	110	After Treatment, 148
Variant Strain Present	*pre*	*70*	*70*	*70;215*	*70;215;41*	*215;41*	*215;41*

Figure 4.3 HIV variant strains appearing during the course of infection and treatment.

BEST COPY AVAILABLE

FOR YOUR INFORMATION
The Evolution of RNA Viruses

Many viruses have genomes of RNA. Some are important human pathogens that have significant public health and economic impacts: common cold, flu, human immunodeficiency virus (HIV), and measles, among others. Many examples are also found among plant viruses, several of which have major ecological and economic impacts in agriculture and forestry.

As in the case of bacteria, any virus mutates as it replicates. A distinctive feature of RNA viruses, however, is their very high rate of mutation (10^{-3}–10^{-4} per base pair per replication, a rate 100,000 times greater than for cells). As discussed in Exercise 4.1, the frequency with which mutations arise in any population is determined in part by how often the organisms reproduce. Viruses replicate millions of times each day, so random mutations are constantly arising.

Another factor contributing to high mutation rates of viruses is the relative infidelity of viral nucleic acid replication. The replicase enzyme that copies RNA genomes occasionally makes random errors, inserting the incorrect monomer (for example, A opposite G, or U opposite C). The polymerases that copy DNA also make errors, but unlike RNA viruses, cells have evolved a molecular quality-control, proofreading mechanism that is able to correct mistakes most of the time and thus keep DNA's mutation rate low. The relative "sloppiness" of RNA replication enhances the rate of new mutations. It has been estimated that among the HIV viral population in a single infected person, each of the virus's 10,000 RNA bases is mutated more than 10,000 times *each day*.

As a result of their high mutation rates, any RNA virus population will contain a high level of sequence variation. Thus the "genome" of a population of RNA viruses is not a single unique sequence, but rather a population of many related variants. Because mutations occur randomly, some viral genomes might have escaped mutation entirely, while others may have many mutations. Most of the mutations are harmful, and those viruses will not survive. Some genome changes, however, will by chance confer an advantage for viral propagation or other aspects of viral behavior. Because the evolutionary principle of selection applies to viruses as well as to free-living organisms like bacteria, we can make certain predictions about the general course of virus evolution in the face of medical attempts to thwart them with antiviral drugs.

Each of these viral variants was found to contain one or more mutations in the RNA gene that encodes reverse transcriptase. Rather than impairing the enzyme's function, these particular mutations rendered the enzyme less susceptible to the AZT drug, thus conferring a degree of drug resistance on the variant virus. Variants that have more than one such mutation in the same RT gene are designated with multiple numbers (for example, variant *70;215* has two different mutations). Each number in the designation refers to a mutation in RNA that affects the amino acid at that position number in the RT protein. Variant *70;215*, for example, has two mutations in its RNA, one affecting amino acid 70 and the other affecting amino acid 215 of the RT protein.

You can see in Figure 4.3 that the virus population changed across time in this patient. The change in the viral population is a good example of evolution in a microcosm—in many ways similar to the test tube evolution of RNA that you explored in Activity 3.

You can better understand the historical pathway of HIV evolution within this patient by identifying the ancestor-descendant relationships among the different variant strains. Figure 4.4 is a branching diagram (a *phylogeny*) that allows you to summarize these relationships.

Procedure

1. Use the data in Figure 4.3 to fill in the phylogeny diagram. In the square brackets [ ], write in the variant strain that was present at a particular time (for example, *215;41*).

2. In the parentheses ( ), write the designation number of the particular new mutation (for example, *215*) that arose in the population at the indicated point in time.

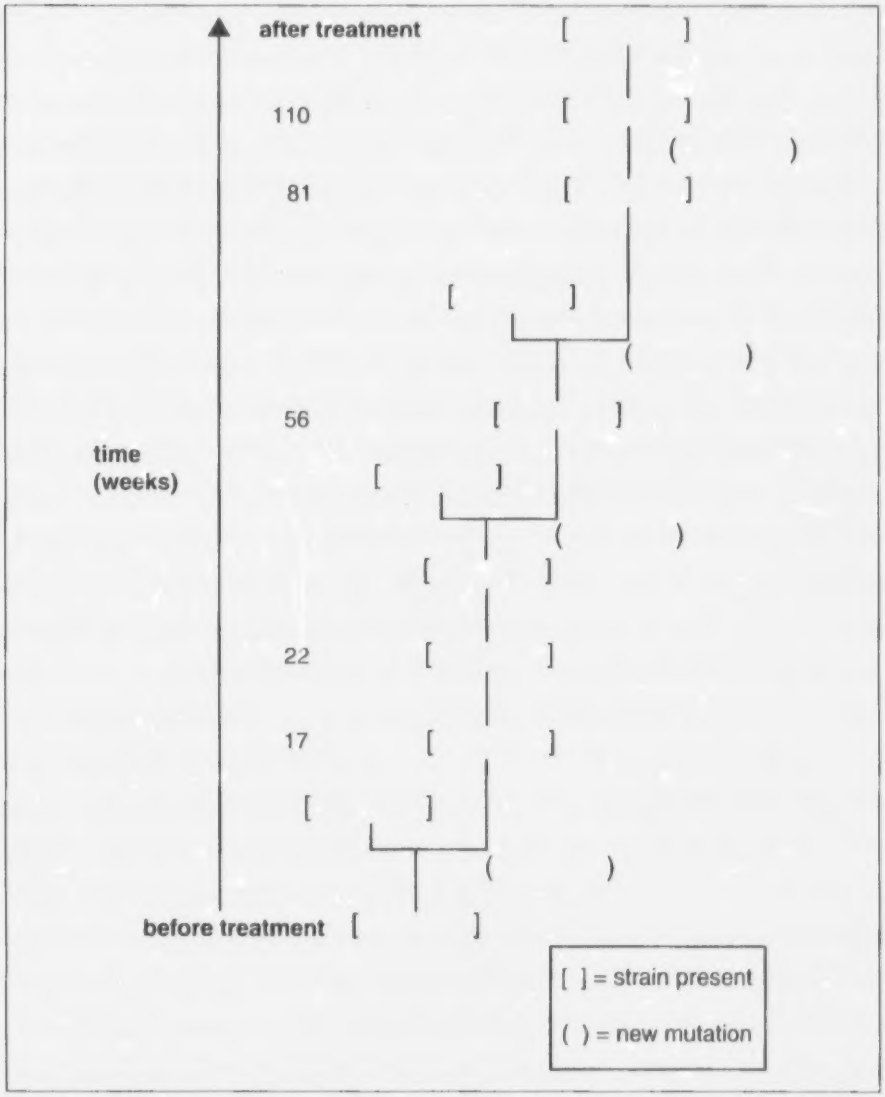


Figure 4.4 A phylogeny of ancestor-descendant relationships showing an evolution of HIV in a single patient.

Challenge Questions

1. The viral population in this patient changed across time. Explain this change in evolutionary terms. What role did AZT play in shaping the course of this change?

2. What do you think happened in the branches (lineages) that appear to terminate during the course of the infection?

3. Which combination of mutations appears to confer the most AZT resistance on the virus during the course of this study? Explain the reason for your choice.

4. What mutational mechanism can account for the eventual loss of the *70* mutation in the surviving virus lineage?

5. The emergence of strain *215;41* sometime between weeks 81 and 110 of therapy, and its persistence after the end of therapy, implies that the rate of change in the viral population slowed when therapy stopped. Explain this slowing in evolutionary terms. What might you expect to happen if therapy with AZT alone were reinstituted? What can you conclude about the long-term effectiveness of treatment with only a single drug?

FOR YOUR INFORMATION
Combating Viral Disease

Antibiotics kill bacteria and other cellular microbes by interfering with cellular structure or function. Because cells are complex structures, there are many points at which they can be vulnerable to different antibiotics. You have seen how the aminoglycosides interfere specifically with bacterial protein synthesis by binding to bacterial rRNA. Penicillin, in contrast, impairs synthesis of the bacterial cell wall. (Eukaryotic cells are unaffected because they lack such walls.) Still other classes of antibiotic interfere with functions of the bacterial cell membrane, such as ion transport.

Because viruses are noncellular structures (they must parasitize living cells to reproduce), they are not affected by antibiotics. Taking an antibiotic for your head cold or flu is not only a waste of time and money, it needlessly exposes your normal bacterial flora to a selective pressure that promotes antibiotic resistance. A variety of other chemical agents known as antivirals have been developed that target different steps in the viral reproduction cycle. Antivirals typically interfere either with the entry of the virus into the cell, the machinery of viral genome replication, or the assembly of progeny virus particles. In the case of HIV I, the viral cause of AIDS, antivirals in current use fall into two main categories: those that interfere with key viral enzymes needed for viral RNA replication (for example, inhibitors of reverse transcriptase or integrase), and those that impair the maturation of viral proteins and the assembly of progeny virus particles (for example, protease inhibitors).

The available antiviral agents have not yet been able to cure AIDS. However, recent treatment protocols that are based on evolutionary principles have shown considerable benefit in reducing the number of virus particles in HIV-infected individuals and in delaying the onset of immune system collapse and full-blown AIDS. These evolution-based protocols share the following features:

- They employ two or more antivirals administered concurrently.
- Each antiviral in the mix targets a different viral function or gene product (for example, an RT inhibitor combined with a protease inhibitor).
- Treatment is started as soon as possible following initial exposure to the virus.

6. After reading the FYI essays titled *The Evolution of RNA Viruses* and *Combating Viral Disease*, answer the following questions:

   a. Antiviral treatment protocols commonly use concurrent administration of two or more antiviral drugs, with each drug targeting a different essential viral function or gene product. Explain the evolutionary rationale for this combination treatment.

   b. Explain the evolutionary rationale for starting treatment with antiviral drugs as soon as possible following exposure to the virus.

Exercise 4.5: Evolution in a Larger Population

You have seen how, even in the microcosm of a single patient, viruses can evolve important new properties such as drug resistance in response to "environmental" factors. Here you will examine the relatedness among several strains drawn from the wider population of HIV viruses that infect humans worldwide.

The following partial nucleotide sequence runs from base 2726 to base 2750 of the HIV gene encoding reverse transcriptase, the viral replicating enzyme dealt with in Exercise 4.4. You can view the entire 10,000+ bases of the HIV genome, along with its known mutations, by checking out the HIV Sequence Database on the Internet at http://hiv-web.lanl.gov/.

base 2726 base 2750
...CCATAAAGAAAAAAGACAGTACTAA...

Note that the bases shown are the proviral DNA form of the virus isolated from the human genome.

The reverse transcriptase gene has been sequenced for several different strains of HIV from several different patients worldwide. Many mutations have been docu-

Sequence		Base Position					
		2726	2735	2740	2746	2749	2750
		C	A	G	A	A	A
wild type	1	T	A	G	A	A	A
mutant sequence	2	C	A	G	G	A	A
mutant sequence	3	T	G	A	A	A	A
mutant sequence	4	C	A	G	G	G	G
mutant sequence	5	T	G	G	A	A	A
mutant sequence	6	C	A	G	G	G	A

*Read sequences from left to right.

Figure 4.5 Wild type and mutant bases.

mented in the gene, and those of particular interest are ones that confer enhanced virulence and/or drug resistance to the virus.

Figure 4.5 focuses on particular base positions within the above sequence, and all are known positions of mutations that confer resistance to the antiviral drug AZT. The six different mutant sequences shown were isolated from six different patients who had undergone therapy with AZT. A "wild type" sequence, from an HIV strain that had not been exposed to AZT, is also shown.

Procedure

Arrange the sequences in Figure 4.5 into the phylogeny shown in Figure 4.6, which depicts a plausible ancestor-descendant relationship among these sequences. To do this, you will use the method of parsimony. Parsimony is a decision-making approach in the field of taxonomy that applies the following assumption: *Those sequences that are most closely related to one another will differ by the smallest number of base changes.* For example, sequences that differ from each other at only one base position are presumed to be more closely related than sequences that differ at more than one position.

1. Consider which of the sequences is likely to be the most ancestral. Write this sequence in the square brackets [ ] at the bottom of the phylogeny.

2. Examine the other sequences and apply the parsimony principle to find the "family" relationships that emerged from this ancestral strain. Fill in the appropriate sequences in the square brackets [ ] at the top of the phylogeny.

3. In the parentheses ( ), fill in the specific base-change mutation that occurred and caused divergence of sequences in that branch of the phylogeny (for example, C to T).

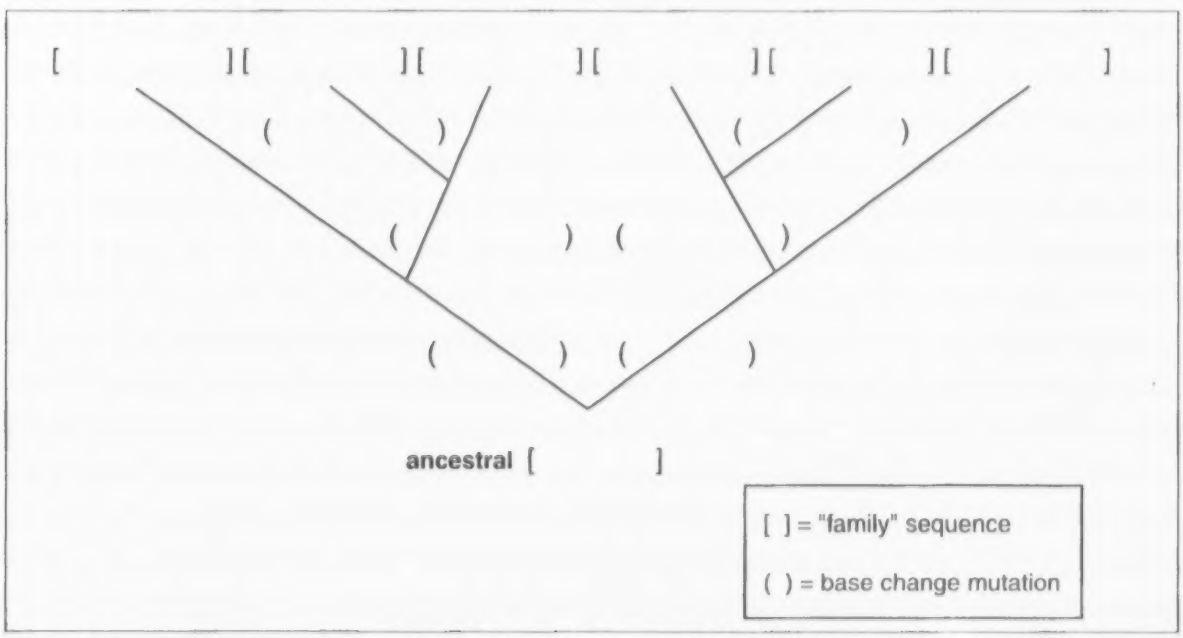


Figure 4.6 A possible phylogeny of ancestor-descendant relationships among several HIV strains.

Challenge Questions

1. As you can see from Figure 4.6, two major lineages (branches) appear to have diverged early from the ancestral sequence. Which newly acquired base at which position is shared by all members of the major lineage on the left? on the right?

2. Which two sequences are more closely related to the ancestral sequence? Which two are more divergent from the ancestral? On what basis do you make these inferences?

3. Recall that these mutations are in the gene encoding the viral reverse transcriptase enzyme, and each of them confers a degree of resistance to the antiviral drug AZT.

   a. Which of the viral strains in your phylogeny have two or more resistance mutations in the reverse transcriptase gene?

   b. Based on your results from Exercise 4.4, what might you predict about the relative drug resistance of the multiple- versus the single-mutant strains of the virus?

FOR YOUR INFORMATION
The Continuity of Life

Only one-half billion years after the earth formed, early cells similar to modern cyanobacteria were living in large, domelike colonies. Today these colonies exist in fossilized form as stromatolites. These oldest of solid clues to the ancient origins of life show us that as long ago as 3.5 billion years, cell structure and biochemistry quite similar to modern life already existed. But how do we observe evidence of the world that existed before 3.5 billion years ago? This was a time before cells, a time when some type of ancient molecule was first carrying out essential activities, such a replication, that characterize life.

Among our best evidence for events on the ancient earth are fossils. But molecules do not make good fossils; they are too small and too fragile. How, then, can we examine that early world? Fortunately, the molecules of our modern world offer a glimpse of the shadows of past events. This connection exists because life involves continuity: Despite great changes that occurred accross time, life and its molecules have existed continuously since its origin almost 4 billion years ago from a single common ancestor. We know this is the case because the molecular machinery of all living cells—from bread mold to bald eagles—is fundamentally similar.

When the base sequences of ribosomal RNAs from many different organisms are compared, for example, the sequences are found to be remarkably similar, although not identical. The similarities immediately suggest that all organisms are related and derive from a single common ancestor. The differences presumably reflect the fact that mutations are constantly arising at random across time in DNA and RNA. Only a few mutations have had a chance to occur in the short time since the organisms diverged from their common ancestor. As a result, closely related organisms have closely related ribosomal RNA sequences, while distantly related organisms have correspondingly more divergent ribosomal RNA sequences. Knowing the degree of relationship between organisms makes it possible to construct a "tree of life" or phylogeny showing the lines of descent from one species to the next.

FOR YOUR INFORMATION (continued)

One of the great surprises of this phylogenetic analysis was Carl Woese's discovery that all living organisms can be divided into three great domains. These are the *eukaryotes* (cells with nuclei), the *eubacteria* (true bacteria), and the *archaebacteria* (ancient bacteria that in some respects appear to be more closely related to eukaryotes than to true bacteria). Both groups of bacteria lack nuclei and are also referred to as *prokaryotes*. Remarkable conservation of ribosomal RNA sequence is found among organisms as diverse as the bacterium *Escherichia coli* (which populates our intestines), the archaebacterium *Halobacterium halobium* (which colors the salt flats of San Francisco Bay red), and the unicellular eukaryote *Trypanosoma brucei* (the cause of African sleeping sickness). This molecular similarity strongly supports the idea of an underlying relatedness among these organisms. The fact that ribosomal RNA has changed relatively little across time supports the idea that this molecule plays a key role in protein synthesis, almost certainly as the catalyst of peptide bond formation. (RNA's ability to carry out this reaction also reinforces the notion that the very first ribosomes consisted simply of a catalytic RNA able to join amino acids, and that modern ribosomal RNA has retained these catalytic functions despite assistance from more than 50 ribosomal proteins added later). Recall that certain fungal-derived antibiotics such as streptomycin and neomycin exploit both the indispensability of rRNA and the subtle sequence differences between prokaryotic and eukaryotic rRNAs to target and kill bacteria (recall Exercises 4.2 and 4.3).

Biologist A.G. Cairns-Smith used the following analogy to describe the continuity of modern and ancient forms of life: "None of the fibers in a rope has to stretch from one end to the other, so long as they are sufficiently intertwined to hold together sideways." For life, the genetic information passed repeatedly from one generation to another acts like the fibers in a rope and provides continuity across time. No single gene sequence stretches unaltered all the way across the billions of years, but life's collection of continuously replicated, related genetic information does form an unbroken chain from the early earth to the present.

Bringing RNA into View: RNA and Its Roles in Biology

Contains	■	faculty background materials, including discussion of many new discoveries in the field
	■	classroom activities, with faculty pages, copymasters, templates, and student pages for four activities
Designed to	■	provide materials to teach about the nature and methods of studying RNA and for use in professional development for instructors
Developed by	■	BSCS (Biological Sciences Curriculum Study)
Supported by	■	the National Science Foundation under grant No. 9652921

FREE monograph for the college biology classroom

INNOVATIVE SCIENCE EDUCATION SINCE 1958

5415 Mark Dabling Blvd.
Colorado Springs, CO 80918-3842
(719) 531-5550 • Fax (719) 531-9104
www.bscs.org

BEST COPY AVAILABLE

SE 0A1476

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Bringing RNA into View, RNA and Its Roles in Biology	
Author(s): BSCS	
Corporate Source:	Publication Date: 2000

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: Barbara Perrin	Printed Name/Position/Title: Barbara Perrin	Production Manager
Organization/Address: BSCS 5415 Mark Dabling Blvd Colorado Spgs CO 80918	Telephone: (719) 531-5[illegible]	FAX: (719) 531-7[illegible]
	E-Mail Address: bperrin@bscs.org	Date: 8/23/02

AUG-23-2002 FRI 11:48 AM BSCS FAX NO. 719 5319104 P. 01

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfacility.org